IMPACT OF WORKING MEMORY BURDEN AND CONTEXTUALIZATION ON COGNITIVE COMPLEXITY

A Thesis
Presented to
The Academic Faculty

By

Kristin M. Morrison

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science in the
School of Psychology

Georgia Institute of Technology

May, 2013

Impact of Working Memory Burden and Contextualization
On Cognitive Complexity

Approved by:

Dr. Susan Embretson, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Eric Schumacher
School of Psychology
Georgia Institute of Technology

Dr. Lawrence James
School of Psychology
Georgia Institute of Technology

Date Approved: April 4, 2013

ACKNOWLEDGEMENTS

I would like to thank my wonderful parents, Todd and Brenda, for being my rock while I make my way through this journey known as grad school and life. Their continual support and ever-lasting love are constant reminders that I can aspire to anything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Contextualization is often added to mathematical achievement items to place targeted mathematical operations in a real world context or in combinations with other mathematical skills. Such items may have unintended sources of difficulty, such as greater cognitive complexity than specified in the test blueprint. These types of items are being introduced to achievement exams through assessment programs such as SBAC and PARCC. Cognitive models have been created to assess sources of cognitive complexity in mathematics items, including a global model (Embretson & Daniel, 2008) and an adapted model (Lutz, Embretson, & Poggio, 2010). The current study proposes a new cognitive model structured around sources of working memory burden with an emphasis on contextualization. Full-information item response (IRT) models were applied to a state accountability test of mathematics achievement in middle school to examine impact on psychometric properties related to burden on working memory.

CHAPTER 1
INTRODUCTION

Many different, basic skills are needed in academic settings, such as reading, writing, and mathematics and most of these skills are taught separately from each other. However, many different areas require the combined application of these skills. For example, one might learn how to read and write in an English class but then must apply these skills to a Science assignment. Teaching these basic skills in the context of other areas is considered contextualization (Perin, 2011). Contextualization can be observed in the given example or when mathematical items require more than basic mathematic skills to solve. Examples include mathematical word problems or problems that include information not necessary to find a solution to the problem. Higher contextualization of mathematics items can lead to increased burden on working memory, inclusion of less content-specific information, and increased item difficulty.

The primary goal of this research is to model cognitive complexity using full-information item response models using an assortment of variables to represent burden placed on working memory and contextualization. Sections of this paper include: a background on working memory, item variation, contextualization, and potential variables; a breakdown of the method for the research; presentation of full-information models to be examined; results from the full-information models; and a discussion of pilot results.

CHAPTER 2
BACKGROUND

Working memory capacity has been shown to be related to both item complexity and general fluid intelligence (Ackerman, Beier, & Boyle, 2005; Ashcraft & Krause, 2007; Unsworth & Engle, 2007). Multiple assessment programs are currently in their developmental stages, set to be structured around a set of Common Core State Standards. These programs are meant to provide help to teachers, students, and parents in an attempt to promote academic achievement. However, they may introduce assessments that are heavily contextualized, which may adversely impact achievement, both on scoring and interpretation of scores. These achievement assessments consist of different item types, or variation, which can be measured through various variables measuring contextualization and burden on working memory. Potential variables can then be combined to model cognitive complexity.

2.1 Working Memory, Intelligence, and Mathematics

Working memory (WM) has been defined in multiple ways but each version is related. Swanson and Beebe-Frankenberger (2004) defined working memory as “a processing resource of limited capacity involved in the preservation of information while simultaneously processing the same or other information” (p. 471-472). Oberauer (2002) described working memory as “a system for simultaneous storage and processing of information” (p. 411). Cowan (2000) and Schacter, Wagner, and Buckner (2000) also defined working memory in similar terms. All definitions include the maintenance of memory representations while simultaneously dealing with distractions, attention shifts, and concurrent processing. Unsworth and Engle (2007) distinguish between two primary functions of working memory, maintenance and retrieval. Maintenance allows new information to remain active. Since there is a limit to the amount of information that can be actively maintained, retrieval is necessary when information is needed in the presence of relevant and irrelevant information. This retrieval is cue-dependent; it requires the differentiation of relevant and irrelevant information given a combination of cues related to information being maintained. Thus, working memory is not one thing. It is a combination of: the ability to maintain information in primary memory (PM), control of attention, and retrieval of information from secondary memory (SM).

Working memory capacity (WMC) is the difference between individuals in the ability to actively maintain relevant information and suppress irrelevant information (Engle, 2002). These differences occur due to individual's ability to maintain information in PM and use cues to retrieve information from SM (Unsworth & Engle, 2007). Differences in WMC are related to differences in controlled attention, or commonly known as the central executive (Engle, Tuholski, Laughlin, & Conway, 1999). WM and WMC have been shown to be related to fluid general intelligence (*Gf*), short-term memory (STM) capacity, and higher order cognitive abilities (Engle et al., 1999; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Ackerman et al., 2005; Unsworth & Engle, 2007).

WM may become overloaded for a variety of reasons. An individual might try to maintain more information in PM. However, due to a capacity limit on memory (Cowan, 2000), information is redirected to SM. Whenever information from SM is needed to complement the information in PM, a controlled search and retrieval process occurs. As demand increases for maintenance and retrieval of information in PM and SM, an individual's WM is taxed. Differences in WM occur because of the inefficiency of the controlled attention aspect to alternate between PM and SM, rotating between the information being stored in these systems. Burden on WM can be increased due to the inability to effectively monitor the relations and processes in WM or the amount/type of information trying to be stored in WM; this might occur in a variety of situations and contexts, such as mathematics.

Floyd, Evans, and McGrew (2003) found mathematics performance to highly correlate with working memory and *Gf*. The relationship between WM and mathematics relied heavily on the controlled aspect of WM (i.e., the central executive). Ashcraft and Krause (2007) found a positive relationship between the complexity of mathematical items and the demand imposed on working memory. As an item's complexity increased, an individual had to rely more on WM. This stronger reliance increased the burden and demand on WM. If the controlled aspect of WM is unable to efficiently consolidate information in PM and SM, the burden placed on WM may increase. Although an individual's WMC may influence mathematics ability, in general, the burden placed on WM may also affect mathematics performance.

2.2 Assessment Programs

Bennett (2010) introduces what is called "an innovative K-12 assessment system that integrates learning-sciences theory with content standards". The project, the Cognitively Based Assessment of, for, and as Learning (CBAL), is intended to measure and promote student achievement, both on the student and teacher fronts, with formative and summative assessments. In his paper, Bennett describes the beneficial consequences of the program, along with a few of the unintended consequences. These unintended consequences include “teaching to the test” and creation of tests that are biased towards certain formats and strategies.

However, another possible downfall to the program was discussed by Embretson (2010). A possible unintended, detrimental consequence is that the model highly emphasizes the use of critical thinking skills in both assessments and instruction. This overemphasis may have a negative impact on student achievement instead of the intended positive impact. By focusing intensely on the integration of critical thinking skills and mathematical problems, there may be several negative consequences, such as: “1) less optimal psychometric properties of the assessments, 2) adverse impact on several aspects of validity, and 3) decreased emphasis on crucial mathematical skills in instruction and assessment” (Embretson, 2010).

In the past, states within the United States evaluated mathematics achievement differently, partly due to differing mathematics curriculum among the states. Now, the country wants to improve mathematics achievement, which calls for a more focused and coherent mathematics curriculum. The Common Core State Standards (CCSS) for mathematics are a potential answer to this problem (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010); these are internationally benchmarked, focused, and coherent standards. These Standards explain what students should know, understand, and be able to do within the area of mathematics in order to succeed in higher learning. Each grade level will have a set of CCSS to consult for the development of mathematical items. Different programs are being developed in order to incorporate these CCSS into mathematics assessment exams.

One program which incorporates the CCSS is the Smarter Balanced Assessment Consortium (SBAC) (Smarter Balanced Assessment Consortium: Executive Summary, n.d.). The program hopes to provide students with educational experience and opportunities that enable them for success in postsecondary education or a career upon graduation from high school. SBAC attempts to create an over-arching state assessment system using the CCSS as a foundation. This comprehensive system calls for a use of several different items types and performance events for measurement. The system will also be able to provide accurate measurement for all students regardless of individual differences. A goal of SBAC is to provide resources/tools to schools and teachers that improve instruction, helping students succeed, through the use of the assessment system. Evidence-centered design is a concept used by SBAC to provide a framework for creating its system. Employment of this concept helps in analyzing the CCSS, which influences the development of SBAC's content specifications, leading to item and task specifications, culminating in a sample of items and performance tasks.

Items and tasks developed fall into six categories: selected response, constructed response, extending response, performance tasks, technology-enabled, and technology-enhanced (SBAC: Introduction, 2012). These later categories allow the items to be administered on computers. Performance tasks must include multiple ideas, one of which is to "reflect a real-world task and/or scenario-based problem" (SBAC: Performance Task Specifications, 2012). All items and tasks have general specifications that must be followed, along with grade-specific and content-specific specifications as well. To achieve its goals of providing teachers with more information, accurate measurement, and more efficient assessments, SBAC intends to capitalize on the use of computerized adaptive testing (CAT), which allows for more innovative and real-world item types. By working in conjunction with Partnership for Assessment of Readiness for College and Careers, SBAC hopes to help states transition to CAT successfully.

Partnership for Assessment of Readiness for College and Careers (PARCC) includes 23 states working together to develop improved, and common, assessments in both English and Mathematics (PARCC, 2012). Assessments developed by PARCC are to be founded on the CCSS, just like SBAC. They also share the goal of providing students with the necessary skills and knowledge to allow them to be college and career ready upon graduation from high school. Beneficiaries of the program include: students, teachers, parents, states, and the nation (PARCC, 2012). PARCC's assessments will occur closer to presentation of the material, providing teachers with key information that may help them adjust instruction. The assessments will provide information which can be easily compared across states, provide information to students, and they will be computer based for efficiency and better technological use in assessments.

Each grade level will have PARCC assessments, composing of both summative and non-summative assessment components. These assessments, as with SBAC, will include a variety of item types, such as constructed response, performance tasks, and selected response. Also in accordance with SBAC, performance tasks will provide students with contextual items which require mathematical reasoning to find a solution (PARCC: 3-8 Assessments, 2012). Frameworks will be used in assessment design to bridge the gap between the CCSS and PARCC assessments. Item types will be informed using these frameworks. PARCC and SBAC both hope to create an assessment system to benefit teachers and students at different points along the education route using the

Common Core State Standards. Both programs were awarded grants from the U.S. Department of Education through the Race to the Top Assessment Program (Race To The Top, 2012).As with CBAL, these programs may have downfalls not expected by educators. These downfalls include teaching to the test and format/strategy biases. Inclusion of multiple different item formats and real-world scenario based problems (contextualization) could also be potential problems associated with the programs.

Heavy contextualization of mathematics items poses a challenge for reporting information in a useful way to students and teachers. Contextualization may also impact a student's WM; an increase in the burden imposed on WM could pose other problems. One byproduct to an increase in burden on WM is an unintentional increase in the difficulty of an item. As items become more contextualized, they may inadvertently become more difficult. Another byproduct of an increase in the burden on WM is that assessments may no longer be measuring content-specific information. The assessments may be measuring basic skills, such as reading and writing. As mathematics items are being heavily contextualized, more of these basic skills may be required. This may be reflected in an increase in the burden placed on working memory, and the items may no longer be measuring mathematics ability alone but rather a conglomeration of skills and abilities. In fact, the items may be better indicators of general fluid intelligence (*Gf*). The model proposed in this paper focuses on the sources of burden on WM, specifically which contextual variables better measure the extent of contextualization in mathematical assessment items and other aspects that increase the demand on WM.

2.3 Item Variation

Assessing variation within assessment items is a subject of interest in both past and recent literature (Birenbaum, Tatsuoka, & Gurtvirtz, 1992; Daniel & Embretson, 2010; Gorin, 2005; Gorin& Embretson, 2006). Item response theory (IRT) has become mainstream as a basis for determining psychometric properties. IRT has proven to be an improvement over simple classical test theory (CTT) statistics, providing refined trait level and item property estimates (Embretson and Reise, 2000). This approach to calibration of psychometric properties by IRT also has a broader scope in application than CTT. Certain tests, such as computerized adaptive testing (CAT), heavily rely on IRT (Embretson & Reise, 2000). Item difficulty has been shown to be related to various components in both mathematical and verbal assessments (Bejar, 2010; Daniel & Embretson, 2010; Embretson & Daniel, 2008; Gorin & Embretson, 2006). Gierl, Leighton, and Hunka (2000) noted several advantages associated with understanding the cognitive complexity of achievement exams. These advantages include increased understanding of test scores, increased evidence for construct validity, explication of variables for item design, and unification of theories of achievement, ability, and instruction.

Cognitive theory can be applied to item development principles through use of cognitive design approaches (Embretson, 1999). By scoring stimulus features of items, mathematical models of item difficulty can be developed. These stimulus features can predict item difficulty as well as pinpoint different sources of complexity within the item. Embretson (1999) listed four advantages of cognitive design approaches over traditional item development methods. These advantages include better prediction of item properties, item-level construct validity is assessed, items can be developed for specific properties, and cognitive design approaches allow for item generation.

Embretson and Daniel (2008) applied the linear logistic test model (LLTM) to mathematical problem solving items from the Quantitative Reasoning section of the Graduate Record Examination (GRE). Results of this application support the notion of a processing model of item difficulty. The model included five stages of processing: encoding, integration, solution planning, solution execution, and decision processing. Variables included in these stages were found to impact item difficulty. This model has been able to predict item difficulties on broad tests of mathematical reasoning but has also been used to generate items.

An adaptation of this model is currently being used to predict mathematical item difficulty on middle-school standards-based mathematics assessment exams (Lutz, Embretson, & Poggio, 2010). Items included on these assessment exams are created from a blueprint of four standards: Number and Computation, Algebra, Geometry, and Data. Within these four standards are ten benchmarks and a total of 25 indicators. Each indicator can be represented by multiple items and each item within the indicator can vary in difficulty. The current model consists of five major cognitive components, each consisting of different attributes that further define the component. These components are structured around the processing stages required to get a solution. Although the Embretson and Daniel (2008) model and the Lutz et al. (2010) adapted model had moderate success in predicting item difficulty, they did not focus directly on the burden placed on working memory. However, working memory burden, when excessive, may impact examinee's capability to perform mathematical operations. An alternative, resource-based cognitive model was created to focus on sources of burden on working memory (i.e., focuses on the cognitive demands imposed by an item).

2.4 Contextual Variables

Contextualization has many different facets associated with the concept. As previously defined, the overarching idea of contextualization is defined as teaching basic skills, such as reading and writing, within the context of other content areas (Perin, 2011). Perin (2011) showed that this 'umbrella' idea can be further broken down into two forms: contextualized basic skills instruction and integrated basic skills instruction. Contextualized basic skills instruction involves the teaching of academic skills within a specific subject matter to which the skills need to be applied. For example, an English instructor might use information regarding a science topic, such as the water cycle, to teach how to write a cause and effect essay. Integrated basic skills instruction incorporates basic skills instruction (such as reading instruction) into instruction of a specific discipline; the inclusion of these skills increase critical thinking of the subject matter. Using this instruction, a science instructor might teach the students how to write a cause and effect essay to explain the water cycle. Both forms of contextualization are similar with slight variations in how and why academic skills are being used in domain-specific content areas. In the realm of mathematics, both types of contextualization could be observed through the presentation of word problems. The inclusion of contextualization in mathematical problems adds to the cognitive complexity already present due to the mathematics aspect of the problem. This additional, and possibly irrelevant, cognitive complexity may limit an individual's ability to perform mathematical operations. Working memory is central to these processes and when burden on it increases, performance is affected.

Different methods exist for measuring contextualization. These methods range from text comprehension, Latent Semantic Analysis (LSA), Computerized Propositional Idea Density Rater (CPIDR 3, version three), and propositional analysis. All but CPIDR 3 were examined for inclusion into the resource-based cognitive model. A comparison was performed between the results from CPIDR3 and the propositional analysis conducted.

2.4.1 Text comprehension

The amount of contextualization within an item can be measured through text comprehension, which can be determined using several different variables. Text comprehension plays a role in the demand placed on working memory when solving an item. As text comprehension becomes more difficult, more demand is placed on working memory to solve the item; therefore, the item becomes more difficult. Simple encoding variables, both contextual (Dark & Benbow, 1991; Jones & Anderson, 1987; Kintsch & van Dijk, 1978) and mathematical (Ashcraft & Battaglia, 1978; Ashcraft & Krause, 2007; Widaman, Geary, Cormier, & Little, 1989), influence the difficulty of an item; contextual variables are simply a count of words in the item and mathematical variables are a count of the mathematical terms or operators within an item. Researchers have found that as the length of information to be remembered increased, the cognitive load, or burden, on working memory increased because there is more information to sift through in order to determine what to use and how to use the information. With this increase, performance decreased. Previous research has also shown that fewer cognitive resources are usually needed when a text is easier to read (Hitch, Towse, & Hutton, 2001) and when the text has a lower or equal comprehension level of the target population (Ashcraft & Kirk, 2001). The Flesch Reading Ease Test determines the readability of an item's stem; the higher the text scores on this measure, the easier the text is read. An item's reading comprehension level is determined with the Flesch Kincaid Grade Level Test; this measure gives the United States grade level of the text. Preferably, text should be at or below the targeted grade level.

2.4.2 Latent Semantic Analysis

Item difficulty and working memory burden are also influenced by the contextual relatedness of the text. An automated method to analyze text comprehensibility is Latent Semantic Analysis (LSA), which creates a semantic space between words, sentences, and paragraphs (Landauer, 1998; Landauer, Foltz, & Laham, 1998). LSA has been used to determine meaning similarity between words and the consequences of these word similarities in paragraphs (Landauer et al., 1998). Similarities can be determined with five different tests: near neighbors, matrix comparison, sentence comparison, one-to-many comparison, and pairwise comparison; for a complete description of each approach, please see the LSA website. An interesting aspect of LSA is the ability to choose between different LSA semantic spaces for the analysis. For middle-school assessment exams, which are of primary interest in this research, a LSA space of general reading up to 9th grade is most appropriate. For mathematical items, sentence comparison is best suited for measuring contextual relatedness. Correlations are created between each successive pair of sentences. Higher correlations indicate stronger relatedness between the sentence pair. The more related a set of words/sentences are, the easier they are to encode, leading to better performance (Landauer, 1998; Jones & Anderson, 1987).

2.4.3 CPIDR 3

Covington (2007) has created a computer program, called Computerized Propositional Idea Density Rater (CPIDR 3, third major version), to automatically determine the propositional idea density of an English text as part of the CASPR project (Computer Analysis of Speech for Psychological Research). As cited in Covington (2007), propositional idea density "can be approximated by the number of verbs, adjectives, adverbs, prepositions, and conjunctions divided by the total number of words (Snowdon et al., 1996)". CPIDR 3 refines this technique, using a part-of-speech tagger and readjustment rules, to obtain accurate measures. Output from the program provides: the rules that are applied to each word, part-of-speech tags, the specification W and P to indicate which items were counted as words and propositions, respectively, the number of propositions, the number of words, and the idea density. The program also offers a "Speech mode," which will reject most repetitions of propositions but still count them as words.

2.4.4 Propositional analysis

A propositional analysis, as proposed by Kintsch and van Dijk (1978), provides a more detailed outline of text comprehension. The propositional structure of a text has an influence on its readability, which may in turn impact item difficulty; more specifically, the number of modifier propositions may be influential to a text's comprehensibility. Several studies, as cited in Gorin (2005), have shown that "propositionally dense text is difficult to process and integrate for later recall and comprehension (Kintsch, 1994;

Kintsch & Keenan, 1973; Kintsch & van Dijk, 1978)" (p. 353). This increased processing difficulty increases the burden on working memory as individuals attempt to understand the text. Turner and Greene's (1977) technical report was consulted to conduct a propositional analysis. Three types of propositions exist: 1) modifier propositions, which restrict/limit a concept by using another concept, 2) connective propositions, which relate propositions/facts within a text, and 3) predicate propositions, which express actions or ideas. Hypothetically, as the number of propositions increases, so does the amount of cognitive resources necessary for understanding the text, increasing item difficulty as well. Not only does the number of propositions impact item difficulty, but Embretson, Fultz, and Dayl (1989) found that certain densities can significantly impact difficulty as well.

2.5 Potential Cognitive Variables

Variables considered for the model are thought to influence cognitive demand, which usually affects item complexity and difficulty. Cowan (2000) stated there is a capacity limit to working memory, usually considered to be about four chunks of information. A chunk is "a collection of concepts that have strong associations to one another and much weaker associations to other chunks concurrently in use" (Cowan, 2000, p. 89). The more information required to solve an item increases both item complexity and demand of working memory. As these two factors increase, as shown through the use of these cognitive variables, it is speculated that items become more difficult to solve and mathematics performance decreases. As a person's working memory capacity limit is being reached (i.e., demand on working memory is becoming too great), it becomes difficult to retain all the information necessary for solving a problem. This in turn increases the item's difficulty. These potential cognitive variables were divided into five components, retained from earlier models: translation, integration, solution planning, solution execution, and decision processing. All contextualization variables were considered to be a part of the first component, translation.

Equation source (Daniel & Embretson, 2010) and representations (Hegarty & Kozhevnikov, 1999; van Garderen & Montague, 2003) also influence the cognitive load on working memory. Equations and representations may be given, translated, or even generated/visualized, which impacts working memory demand. Given equations and representations would be the easiest source, with increasing difficulty for other sources. One key component of mathematical talent is to be able to translate linguistic information into a mathematical format (Dark & Benbow, 1990); without this talent, exceptional mathematics performance is difficult. The presence and number of subgoals also increases the demand on working memory needed to solve an item (Ashcraft & Kirk, 2001; Tarmizi & Sweller, 1988). Ashcraft and Krause (2007) stated that as the number of subgoals increased, the demand on working memory needed to find a solution also increased. Previous research provided evidence for examining procedural knowledge requirements (Ashcraft & Krause, 2007), semantic memory requirements (Schacter, Wagner, & Buckner, 2000), metacognition processes (Deshler & Lenz, 1989), and mathematical propositions (Dark & Benbow, 1990). Items requiring more computations (Geary, 1993) and involving more operands (Zentall, 1990) increase cognitive load as well. Evidence has been found that as the size of the numbers used within the question grows, item difficulty and working memory demand also increased (Ashcraft & Battaglia, 1978; Ashcraft & Krause, 2007; Widaman, Geary, Cormier, & Little, 1989).

The role of distractors (decision processing) even influences working memory (Birenbaum, Tatsuoka, & Gurtvirtz, 1992; Embretson & Daniel, 2008; Gorin & Embretson, 2006; Haladyna & Downing, 1988; Wakefield, 1958). Relevancy also influences working memory burdens; Cooney and Swanson (1990) stated that the ability to get rid of extraneous information is important. Demand placed on working memory is measured by the presence or amount of these variables within a given item.

As contextualization of mathematics items rapidly becomes a popular idea, unintentional negative side effects may occur, such as increased demand on working memory leading to increased difficulty of items. Therefore, a model showing the difficulty of items and the source of the difficulty will help to show what cognitive components, and hence what attributes, may lead to increased difficulty.

CHAPTER 3
METHOD

This section provides information on the assessment exam, and its examinees, that will be used to develop the cognitive model. IRT estimates were obtained for initial examination of the items composing the exam and to compare against the full-information IRT model. Two full-information models are explicated and then applied to the data.

3.1 Assessment Exam

Development of this cognitive model will use a single 8^{th} grade end-of-the-year mathematics assessment exam from the public school system of a Midwestern state. The exam is composed of four blueprint standards: Number and Computation, Algebra, Geometry, and Data. Within these standards are ten benchmarks: Number and Computation contains three benchmarks, Algebra contains three benchmarks, Geometry contains two benchmarks, and Data contains two benchmarks. A total of 25 subindicators result from these benchmarks. Table 1 provides definitions for these subindicators. The exam is composed of 86 multiple-choice items; the items were broken up into three sections. Each of these sections tested specific principles relating to the indicators, with some overlap between sections.

Table 1. *Blueprint standards, benchmarks, and subindicators with their definition.*

Standard 1: Number and Computation

- Benchmark 1: Number Sense
  - M.8.1.1.K5a – knows and explains what happens to the product or quotient when a positive number is multiplied or divided by a rational number greater than zero and less than one
  - M.8.1.1.K5b – knows and explains what happens to the product or quotient when a positive number is multiplied or divided by a rational number greater than one
  - M.8.1.1.K5c – knows and explains what happens to the product or quotient when a nonzero real number is multiplied or divided by zero
- Benchmark 2: Number Systems and Their Properties
  - M.8.1.2.K2 – identifies all the subsets of the real number system [natural (counting) numbers, whole numbers, integers, rational numbers, irrational numbers] to which a given number belongs
  - M.8.1.2.A1a – generates and/or solves real-world problems with rational numbers using the concepts of these properties to explain reasoning: commutative, associative, distributive, and substitution properties
  - M.8.1.2.A1b – generates and/or solves real-world problems with rational numbers using the concepts of these properties to explain reasoning: identity and inverse properties of addition and multiplications
- Benchmark 4
  - M.8.1.4.K2a – performs and explains these computational procedures with rational numbers: addition, subtraction, multiplication, and division of integers
  - M.8.1.4.K2b – performs and explains these computational procedures with rational numbers: order of operations (evaluates within grouping symbols, evaluates powers to the second or third power, multiplies or divides in order from left to right, then adds or subtracts in order from left to right)
  - M.8.1.4.A1a – generates and/or solves one- and two-step real-world problems using computational procedures and mathematical concepts with rational numbers
  - M.8.1.4.A1b – generates and/or solves one- and two-step real-world problems using computational procedures and mathematical concepts with the irrational number pi as an approximation
  - M.8.1.4.A1c – generates and/or solves one- and two-step real-world problems using computational procedures and mathematical concepts with applications of percents

Standard 2: Algebra

- Benchmark 2: Variable, Equations, and Inequalities
  - M.8.2.2.K3a – solves one- and two-step linear equations in one variable with rational number coefficients and constants intuitively and/or analytically

Table 1 (continued)

		M.8.2.2.A1a – represents real-world problems using variables, symbols, expressions, one- or two-step equations with rational number coefficients and constants
	Benchmark 3: Functions	
		M.8.2.3.A3 – translates between the numerical, tabular, graphical, and symbolic representations of linear relationships with integer coefficients and constants
	Benchmark 4: Models	
		M.8.2.4.A2 – determines if a given graphical, algebraic, or geometric model is an accurate representation of a given real-world situation
Standard 3: Geometry		
	Benchmark 1: Geometric Figures and Their Properties	
		M.8.3.1.K6a – uses the Pythagorean theorem to determine if a triangle is a right triangle
		M.8.3.1.K6b – uses the Pythagorean theorem to find a missing side of a right triangle where the lengths of all three sides are whole numbers
		M.8.3.1.A1a – solves real-world problems by using the properties of corresponding parts of similar and congruent figures
	Benchmark 4: Geometry from an Algebraic Perspective	
		M.8.3.4.K1a – uses the coordinate plane to list several ordered pairs on the graph of a line and find the slope of the line
		M.8.3.4.K1b – uses the coordinate plane to recognize that ordered pairs that lie on the graph of an equation are solutions to that equation
		M.8.3.4.K1c – uses the coordinate plane to recognize that points that do not lie on the graph of an equation are not solutions to that equation
		M.8.3.4.K1d – uses the coordinate plane to determine the length of a side of a figure drawn on a coordinate plane with vertices having the same x- or y-coordinates
Benchmark 4: Data		
	Benchmark 1: Probability	
		M.8.4.1.K3 – finds the probability of a compound event composed of two independent events in an experiment, simulation, or situation
		M.8.4.1.A4a – makes predictions based on the theoretical probability of a simple event in an experiment or simulation
	Benchmark 2: Statistics	
		M.8.4.2.K3 – determines and explains the measures of central tendency (mode, median, mean) for a rational number data set

3.2 Participants

This analysis will use a simple random sample of 2,993 students from the eight-grade. Approximately half (51%) of the students were male. The majority of the students (74%) are White, 11% were Hispanic, 8% were Black, and 1.5% is Native American. The sample will also include students from low income families and those who fell below the curriculum-based performance standards. Data will be stripped of any potential identifiers.

3.3 IRT Estimation

Scored responses from the eight-grade students were used for calibrating the items composing the exam. Four different item-response (IRT) models were estimated using the data: a 1-parameter logistic (1PL) model, 2-parameter logistic (2PL) model, a 3-parameter logistic common lower asymptote (3PLC) model, and 3-parameter logistic (3PL) model. The common lower asymptote was constrained to be equal across items for the 3PLC model. BILOG-MG with the EM algorithm was used for the calibration of parameters for each model (Zimowski, Muraki, Mislevy, & Bock, 2003). Likelihood values and AIC values for the four models can be found in Table A1. BILOG-MG does not report AIC values; Equation 1 was used to calculate the AIC for each model. A classical item analysis was conducted in order to examine the classical test theory (CTT) statistics for all items using all the observations (Table A2). Parameter estimates for the 1PL and 2PL models can be found in Table A3; estimates for the 3PLC and 3PL models can be found in Table A4. The 3PL model best fits the data; however, later full-information models are Rasch-based, the fit of the 1PL and 2PL models were examined.

Neither of these models fit the data as well as the 3PL model, but they are not ill-fitting either; the 2PL model does fit the data better than the 1PL model.

$$AIC = -2\ln L + 2K, \text{ where } K = \text{the number of parameters in the model} \quad (1)$$

Based on responses by the examinees, 70.55% of the exam items were answered correctly; this implies that, on average, the items were easy. Biserial correlations for each item were also examined in order to assess potentially troublesome items. Items with a biserial correlation below 0.30 may be considered inappropriate; these items could be deleted from the analysis. The classical item analysis indicated two items with biserial correlations below this value: Item 4 (r_{b4} = -0.151) and Item 86 (r_{b86} = 0.229). BILOG-MG automatically assumes any item with a biserial correlation less than -0.15 has been miskeyed and omits the item from the 2PL, 3PLC, and 3PL models. Therefore, Item 4 is omitted from all full-information IRT models as well. Item 86 was not excluded but was monitored for problems.

3.4 Cognitive Model Variables.

Based on the literature, previous research, and a pilot study, 68 variables were scored for consideration into the new cognitive scoring model. For a complete list of these variables and their definitions, please see Table B1. Descriptive statistics for all variables can be found in Table B2. An example of an exam item can be found in Table 2, with sample scores for the item in Table 3. Some of the variables scored were used to score a higher-level variable; for example, the variable Number Knowledge receives a score of 1-5 for the highest number knowledge required to solve the item ranging from single-digit (lowest) to fraction/decimal (highest).These higher-level variables were evaluated for inclusion into the model. Many definitions of these variables remained the same as previous literature (Embretson & Daniel, 2008; Lutz et al., 2010); however, some definitions had to be adapted to the context of mathematical achievement items or detecting sources of working memory burden. The scoring process also required a complete propositional analysis to be conducted on the exam. Even though CPIDR 3 has been developed to determine propositional idea density, a traditional propositional analysis was conducted using the original method explicated by Turner and Green (1977). An example of a propositional analysis can be found in Table 4 for the example item given. Duplicate propositions were not included in the analysis (i.e., duplicate propositions were eliminated). Scores for all cognitive variables can be found in Appendix Tables B3 through B16. Zero-order correlations between the psychometric properties of the 1PL and 2PL models with all variables can be found in Table B17.

Table 2. *Example assessment item.*

Maria plans to pay off her credit card balance. She owes a total of $300 and plans to pay $30 each month. The credit card company charges $6 each month in interest, so only $24 is applied toward her balance each month. The table below shows the relationship between the number of months Maria is paying off her balance and how much she still owes.

x	**Number of Months**	0	1	2	5	10
y	**$ Owes**	300	276	252	180	60

Which equation accurately represents Maria's payment plan?

A) $y = 300 - 36x$

B) X $y = 300 - 24x$

C) $y = 300 - 30x$

D) $y = 300 - 6x$

Table 3. *Scores for the example assessment item.*

Translation		
Encoding		
	Total Encoding	117
	Mathematical	44
	Contextual	73
	Stem	89
	Content Words	73
Flesch Reading Ease Test		61.7
Flesch-Kincaid Grade Level Test		5.2
Text Comprehension– LSA (Average)		0.4325
	Comparison One	0.51
	Comparison Two	0.55
	Comparison Three	0.37
	Comparison Four	0.3
Mathematical Propositions		
	Assignment Propositions	0
	Relation Propositions	0
Contextual Propositionalization		
	Total Number of Propositions	30
	Total Proposition Density	0.337
	Number of Predicate Propositions	13
	Predicate Density	0.146
	Number of Modifier Propositions	14
	Modifier Density	0.157
	Number of Connector Propositions	3
	Connector Density	0.033
	Total Number of Unique Arguments	25
	Unique Argument Density	0.281
	Total Number of Arguments	129
	Total Argument Ratio	1.450
	Max. Number of Arguments	13
	Relevant Propositions	11
	Density of Relevant Propositions	0.367
	Relevant Words	47
	Density of Relevant Words	0.528
	Irrelevant Propositions	19
	Density of Irrelevant Propositions	0.633
	Irrelevant Words	42
	Density of Irrelevant Words	0.471

Table 3 (continued)

Encode Diagram	0
Integration	
Translate Word Equation	0
Given Equation – in Stem	0
Generate Eq. or Possible Values	1
Access Equation	0
Auxiliary Diagram	0
Translate Diagram	0
Visualization	0
Semantic Memory	1
Solution Planning	
Presence of Subgoals	1
Number of Subgoals	1
Relative Definition of Variables	0
Solution Execution	
Number Knowledge	3
1. Single-digit	1
2. Double-digit	1
3. Triple-digit	1
4. Four-digit +	0
5. Fraction/Decimal	0
Alt. Procedural Knowledge	2
1. Multiple Steps	1
2. Algebraic Equations	1
3. Mixed fractions	0
Procedural Knowledge	1
1. Integers	1
2. Fractions	0
3. Proportions	0
4. Decimals	0
5. Negative Numbers	0
6. Square Roots	0
Number of Procedures	1
Number of Computations	1
Number of Operands	2
Meta-Cognition Process	1
Decision Processing	
Decision Processing Confirmation	1
Bottom-Up Processing	0

Table 3 (continued)

Top-Down Processing	1
Functional Distractors	3

Table 4. *Propositional analysis for example assessment item.*

P – (REFERENCE, MARIA, HER)
M – (QUALITY OF, CARD, CREDIT)
M – (QUALIFY, BALANCE, (QUALITY OF, CARD, CREDIT))
P – (PAY, O: (QUALIFY, BALANCE, (QUALITY OF, CARD, CREDIT)))
P – (PLAN, A: (REFERENCE, MARIA, HER), O: (PAY, O: (QUALIFY, BALANCE, (QUALITY OF, CARD, CREDITY)))))
P – (REFERENCE, MARIA, SHE)
M – (QUALITY OF, TOTAL, $300)
P – (OWE, A: (REFERENCE, MARIA, SHE), O: (QUALITY OF, TOTAL, $300))
M – (QUALIFY, MONTH, EACH)
P – (PAY, O: $30, G: (QUALIFY, MONTH, EACH))
P – (PLAN, A: $, O: (PAY, O: $30, G: (QUALIFY, MONTH, EACH)))
C – (CONJ: AND, (OWE, A: (REFERENCE, MARIA, SHE), O: (QUALITY OF, TOTAL, $300)), (PLAN, A: $, O: (PAY, O: $30, G: (QUALIFY, MONTH, EACH))))
M – (QUALITY OF, CARD, CREDIT)
M – (QUALITY OF, COMPANY, (QUALITY OF, CARD, CREDIT))
M – (QUALIFY, MONTH, EACH)
M – (QUALITY OF, INTEREST, $6)
M – (QUALITY OF, (QUALITY OF, INTEREST, $6), (QUALIFY, MONTH, EACH))
P – (CHARGE, I: (QUALITY OF, COMPANY, (QUALITY OF, CARD, CREDIT)), O: (QUALITY OF, (QUALITY OF, INTEREST, $6), (QUALIFY, MONTH, EACH)))
M – (QUALIFY, MONTH, EACH)
P – (REFERENCE, MARIA, HER)
M – (QUALIFY, BALANCE, (REFERENCE, MARIA, HER))
P – (APPLY, I: QUALIFY, BALANCE, (REFERENCE, MARIA, HER), O: $24, G: (QUALIFY, MONTH, EACH))
C – (CONDITION: SO, A1: (CHARGE, I: (QUALITY OF, COMPANY, (QUALITY OF, CARD, CREDIT)), O: (QUALITY OF, (QUALITY OF, INTEREST, $6), (QUALIFY, MONTH, EACH))), A2: (APPLY, I: QUALIFY, BALANCE, (REFERENCE, MARIA, HER), O: $24, G: (QUALIFY, MONTH, EACH)))
P – (REFERENCE, MARIA, HER)
P – (REFERENCE, MARIA, SHE)
M – (QUALITY OF, NUMBER, MONTHS)
M – (QUALIFY, BALANCE, HER)
P – (PAY, A: MARIA, O: (QUALIFY, BALANCE, HER), S: (QUALITY OF, NUMBER, MONTHS))
P – (OWE, A: (REFERENCE, MARIA, SHE))

Table 4 (continued)

C – (CONJUCTION: AND, A1: (PAY, A: MARIA, O: (QUALIFY, BALANCE, HER), S: (QUALITY OF, NUMBER, MONTHS)), A2: (OWE, A: (REFERENCE, MARIA, SHE)))
P – (SHOW, I: TABLE, O: RELATIONSHIP, G: (CONJUCTION: AND, A1: (PAY, A: MARIA, O: (QUALIFY, BALANCE, HER), S: (QUALITY OF, NUMBER, MONTHS)), A2: (OWE, A: (REFERENCE, MARIA, SHE))))
M – (QUALIFY, (SHOW, I: TABLE, O: RELATIONSHIP, G: (CONJUCTION: AND, A1: (PAY, A: MARIA, O: (QUALIFY, BALANCE, HER), S: (QUALITY OF, NUMBER, MONTHS)), A2: (OWE, A: (REFERENCE, MARIA, SHE)))), BELOW)
M – (QUALIFY, PLAN, PAYMENT)
M – (QUALIFY, (QUALIFY, PLAN, PAYMENT), MARIA’S)
P – (REPRESENT, O: EQUATION, S: (QUALIFY, (QUALIFY, PLAN, PAYMENT), MARIA’S))
M – (QUALIFY, (REPRESENT, O: EQUATION, S: (QUALIFY, (QUALIFY, PLAN, PAYMENT), MARIA’S)), ACCURATELY)

Results from a pilot study informed the selection of variables included into the model. Twenty cognitive variables, or attributes, were chosen for their potential ability to represent the burden imposed on working memory within the five components. Translation was represented by five attributes: 1) contextual encoding, a count of the total number of words (excluding mathematical terms) in the item; 2) predicate propositions, a count of the number of predicate propositions within the stem of an item; 3) modifier propositions, a count of the number of modifier propositions within the stem of an item; 4) connective propositions, a count of the number of connective propositions within the stem of an item; 5) total argument ratio, a ratio of the number of arguments within an items; stem divided by the total number of words in the stem; and 6) Latent Semantic Analysis (LSA), a comparison of sentence similarity within an item. The Integration component was represented by eight attributes: 1) translate word equation, a binary variable indicating whether the examinee needs to interpret an equation given in context; 2) given equation, a binary variable indicating whether a mathematical equation is given in the stem of the item; 3) generate equation or plausible values, a binary variable indicating whether the examinee must generate or derive equations or plausible values; 4) access equation, a binary variable indicating whether an equation must access an equation from a drop-down box; and 5) auxiliary diagram, a binary variable indicated whether the presented diagram or figure is unnecessary to find a solution; 6) translate diagram, a binary variable indicating whether a diagram or figure presented is necessary to find a solution; 7) visualization, a binary variable indicating whether a representation must be drawn to find a solution; and 8) semantic memory, a count of the total number of unique subindicators within an item.

The solution planning component was comprised of two attributes: 1) number of subgoals, a count of the number of steps needed to solve a problem and 2) relative definition of variables, a binary variable indicating whether one variable is defined by another. Solution execution, the fourth component, was represented by three attributes: 1) number knowledge, an ordinal variable indicating the maximum level of number knowledge needed for an item; 2) procedural knowledge, an ordinal variable indicating the maximum procedural knowledge necessary to solve an item; and 3) number of operands, a counts of the total operands needed to solve an item. The decision processing component only consisted of one attribute, decision processing confirmation, a binary variable indicating whether information found in the distractors is necessary to answer the item. These components and attributes are summarized in Table 5.

Table 5. *Summary of cognitive attributes used to estimate item difficulty.*

Component	Attribute
Translation	
	Mathematical Encoding
	Predicate Propositions
	Modifier Propositions
	Connective Propositions
	Total Argument Ratio
	LSA
Integration	
	Translate Word Equation
	Given Equation
	Generate Equation or Possible Values
	Access Equation
	Auxiliary Diagram
	Translate Diagram
	Visualization
	Semantic Memory
Solution Planning	
	Number of Subgoals
	Relative Definition
Solution Execution	
	Number Knowledge
	Procedural Knowledge
	Number of Operands
Decision Processing	
	Decision Processing Confirmation

Data from CPIDR 3 was not used in the analysis. A comparison was made between the number of propositions found from CPIDR 3 and the traditional propositional analysis as set forth by Turner & Greene (1977). There exists a high correlation between CPIDR 3 and the traditional propositional analysis for total number of propositions excluding repetitions, $r = 0.9484$. However, discrepancies did exist between the two propositional approaches. The traditional approach found an average of 10.4884 propositions while CPIDR 3 found 13.8372 propositions. On average, CPIDR 3 found more propositions than the traditional one ($M = 3.34$), with a range in difference from 2 less propositions to 11 more propositions. For a comparison of propositions found by each approach, along with the difference between proposition counts, refer to Table 6. Also, it was determined that CPIDR 3 was incorrectly classifying propositions. A specific example is when an item refers to a “number set” – “set” was consistently classified as a proposition. It is assumed that the program was classifying “set” as a predicate proposition; however, in the context of the question, “set” is a noun and is not a proposition. Therefore, the traditional propositional analysis was retained for the study, considering the possibility that CPIDR 3 may not accurately propositionalize mathematical items.

Table 6. *Comparison between CPIDR 3 and traditional propositional analysis.*

Item	Traditional	CPIDR 3	Difference
p1i1	9	12	3
p1i2	10	13	3
p1i3	20	26	6
p1i4	13	17	4
p1i5	11	19	8
p1i6	10	10	0
p1i7	2	6	4
p1i8	6	7	1
p1i9	6	9	3
p1i10	4	7	3
p1i11	2	4	2
p1i12	7	10	3
p1i13	9	17	8
p1i14	3	6	3
p1i15	4	8	4
p1i16	12	18	6
p1i17	8	11	3
p1i18	10	16	6
p1i19	7	8	1
p1i20	6	8	2

Table 6 (continued)

p1i21	11	10	-1
p1i22	6	6	0
p1i23	14	15	1
p1i24	14	18	4
p1i25	20	25	5
p1i26	19	24	5
p1i27	30	36	6
p1i28	12	13	1
p1i29	30	34	4
p1i30	13	17	4
p2i1	15	20	5
p2i2	3	5	2
p2i3	17	19	2
p2i4	3	4	1
p2i5	20	20	0
p2i6	12	23	11
p2i7	13	19	6
p2i8	8	12	4
p2i9	8	13	5
p2i10	11	17	6
p2i11	11	13	2
p2i12	12	18	6
p2i13	10	14	4
p2i14	14	21	7
p2i15	17	25	8
p2i16	17	23	6
p2i17	19	20	1
p2i18	8	12	4
p2i19	7	12	5
p2i20	16	20	4
p2i21	15	19	4
p2i22	10	12	2
p2i23	10	11	1
p2i24	15	15	0
p2i25	17	26	9
p2i26	7	10	3
p2i27	8	12	4
p3i1	12	15	3
p3i2	16	14	-2
p3i3	17	22	5

Table 6 (continued)

p3i4	11	17	6
p3i5	13	17	4
p3i6	14	19	5
p3i7	24	27	3
p3i8	3	5	2
p3i9	1	3	2
p3i10	8	12	4
p3i11	2	5	3
p3i12	1	4	3
p3i13	2	5	3
p3i14	3	4	1
p3i15	6	7	1
p3i16	3	5	2
p3i17	8	15	7
p3i18	3	5	2
p3i19	3	5	2
p3i20	5	7	2
p3i21	10	10	0
p3i22	3	6	3
p3i23	4	6	2
p3i24	4	5	1
p3i25	6	9	3
p3i26	9	9	0
p3i27	25	25	0
p3i28	20	22	2
p3i29	15	20	5

3.5 Full-Information IRT Models

Different full information IRT models can be used to assess the impact of stimulus features on item difficulty and item discrimination. One is the linear logistic trait model (LLTM); it is a unidimensional, explanatory model that incorporates item properties into item difficulty prediction (Embretson, 1984). As stated by MacDonald and Kromrey (2011), LLTM is “capable of bridging cognitive processing models and psychometric models" (p. 1). The LLTM can be seen as an extension of the Rasch IRT model (Embretson & Reise, 2000). The model can be represented by Equation 2.

$$P\left(x_{is}=1 \middle| \theta_s, \underline{\tau_k}\right) = \frac{\exp(\theta_s - \beta_i^{'})}{1+\exp(\theta_s - \beta_i^{'})} \tag{2}$$

In Equation 2, $\beta_i^{'}$ is the predicted item difficulty, or the weighted combination of stimulus features representing cognitive complexity. These stimulus features are the nineteen attributes within the five cognitive components. This predicted item difficulty can be estimated from Equation 3.

$$\beta_i^{'} = \sum_k \tau_k q_{ik} + \eta_0 \tag{3}$$

where

q_{ik}= values of stimulus factor k in item i

τ_k = weight of stimulus factor k in item difficulty

η_0= normalization constant (Embretson & Reise, 2000).

LLTM has been shown to have several advantages over simple regression; for further elaboration on these advantages, please see Embretson and Daniel (2008). A likelihood-based fit index can be calculated for each target model; this is comparable in magnitude to a multiple correlation. It gives the proportion of data accounted for the given target model and can be calculated using Equation 4.

$$\Delta^2 = \frac{(-2lnL_{null}) - (-2lnL_{target})}{(-2lnL_{null}) - (-2nL_{saturated})} \tag{4}$$

A second model that can be used to estimate item difficulty based on item properties is a generalization of the 2PL model to include item properties for item difficulty and item discrimination (Embretson, 1999). This model is referred to as the 2PL-constrained model and is represented by Equation 5.

$$P\left(X_{ij} \middle| \theta_j, \boldsymbol{q_i}, \boldsymbol{\eta}, \tau\right) = \frac{\exp\left(\sum_{m=0}^{M} q_{im}\tau_m \left(\theta_j - \sum_{k=0}^{K} q_{ik}\eta_k\right)\right.}{1 + \exp\left(\sum_{m=0}^{M} q_{im}\tau_m \left(\theta_j - \sum_{k=0}^{K} q_{ik}\eta_k\right)\right.} \tag{5}$$

where

q_{im} = values of stimulus factor *m* in item *i* discrimination

q_{ik} = values of stimulus factor *k* in item *i* difficulty

q_0 = unit vector

τ_m = weight of stimulus factor *m* in item *i* discrimination

η_k = weight of stimulus factor *k* in item *i* difficulty (Embretson & Daniel, 2008).

The 2PL-constrained model does not necessarily involve the same explanatory variables for both item discrimination and item difficulty. The likelihood-based fit index for the 2PL-constrained model can be found by Equation 4 as well. These two full-information models were used to predict item difficulty using the stimulus features from the cognitive variables, and the 2PL-constrained was used to add item discrimination to prediction.

CHAPTER 4
RESULTS

4.1 LLTM

A full-information IRT model was applied to the data. The LLTM was used to estimate the impact of the cognitive attributes on item response probabilities. All models were estimated using marginal maximum likelihood estimation using a full information mixed model algorithm. The **Q** matrix for item difficulty was specified by the scored attributes for the items. This **Q** matrix included the above attributes except for LSA. Person parameters were specified as random variables from a standard normal distribution, ~$N(0,1)$.

Three models were specified for the LLTM. Two of these models were the null model and saturated model, needed to compute the likelihood ratio fit statistic. The null model only included the intercept (i.e., all items are equally difficulty) while the saturated model, or full model, is a Rasch model that included unique estimates for each item (i.e., the **Q** matrix is an identity matrix). Fit statistics for these models are shown in Table 7.

Table 7. *Fit statistics*

Model	-2lnL	AIC	Δ
LLTM Null	277,358	277,362	
LLTM	267,119	267,161	0.5065
LLTM Saturated	237,446	237,618	

Estimates from the LLTM null and saturated model can be found in Tables C1 and C2, respectively. Item difficulty was estimated using nineteen of the cognitive attributes (i.e., the target model). Overall, the model had moderate fit, $\Delta = 0.5065$. Stimulus feature weights for this model can be found in Table 8. As can be seen, all of these stimulus features contribute significantly to item difficulty at the 0.05 level. Using these weights and the scores of the items on the cognitive attributes, the difficulty of each item was calculated using Equation 3; these estimates can be found in Table C2. The average item difficulty of these items was -1.3652, indicating that, as a whole, these items can be considered relatively easy. Correlations between the item difficulty estimates from the LLTM target model and the estimates from the 1PL IRT model from BILOG can be found in Table 9.

Table 8. *LLTM item difficulty stimulus feature weights.*

Parameter	Estimate	Standard Error	t-value
Math	-0.0050	0.0005	-10.07*
Predicate Prop	0.0861	0.0025	35.10*
Modifier Prop	-0.0517	0.0021	-24.10*
Connective Prop	0.0721	0.0065	11.04*
Total Argument Ratio	0.6293	0.0179	35.16*
Translate Word Equation	-0.2360	0.0154	-15.30*
Given Equation	-0.4764	0.0196	-24.29*
Generate Eq.	-0.2415	0.0142	-16.99*
Access Eq.	-0.1058	0.0213	-4.98*
Auxiliary Diagram	0.0859	0.0237	3.62*
Translate Diagram	0.7235	0.0166	43.67*
Visualization	0.4019	0.0217	18.54*
Semantic Memory	-0.1248	0.0066	-18.89*
Number of Subgoals	0.3056	0.0097	31.39*
Relative Definition	0.9902	0.0337	29.36*
Number Knowledge	0.0400	0.0035	11.29*
Procedural Knowledge	0.0933	0.0033	28.24*
Number of Operands	0.1741	0.0072	24.22*
Decision Processing Confirmation	0.2227	0.0145	15.33*
u	1.0866	0.0303	35.90*
(Intercept)	-2.1505	0.0311	-69.17*

*p < 0.05

Table 9. *Correlations between target model estimates and BILOG estimates.*

	BILOG – difficulty		BILOG – discrimination
	1PL	2PL	
LLTM - difficulty	0.5103	0.5281	
2PL Constrained – difficulty	0.5029	0.5358	
2PL Constrained - discrimination			0.4393

By examining the stimulus feature weights for the LLTM, attributes can be determined to either increase or decrease item difficulty (i.e., they place more or less burden on working memory). The total argument ratio of an item, the need to translate a diagram, visualize a representation, the number of subgoals, the presence of relative definitions, the number of operands, and decision processing confirmation all significantly increased the difficulty of the item. These attributes would be considered to impose the most burden on working memory when trying to find a solution to the item. Translating a word equation, given an equation, generating an equation, accessing an equation, and the semantic memory requirement of an item decrease the item's difficulty, thus, imposing less of a burden on working memory. The remaining variables, mathematical encoding, number of predicate propositions, number of modifier propositions, number of connective propositions, presence of an auxiliary diagram, number knowledge, and procedural knowledge have no major impact on item difficulty or working memory.

4.2 2PL

The 2PL Constrained model was attempted using the given data. However, the model was not feasible due to empirical model underidentification for marginal maximum likelihood estimation. Therefore, regression estimates are used to determine the impact of the cognitive attributes on item difficulty and item discrimination. Item difficulty was estimated using all the cognitive attributes except for LSA, just like with the LLTM. Item discrimination was estimated using the same cognitive attributes as item difficulty, but replaced the three types of propositions and total argument ratio with one attribute, LSA. A summary of the regression for each model (item difficulty and item discrimination) on to their respective 2PL psychometric properties from BILOG-MG can be found in Table 10. The item difficulty model had an *R* of 0.5563 and the item discrimination model had an *R* of 0.4393.

Table 10. *Regression summary for 2PL IRT parameters vs. cognitive attributes.*

	R	R^2	df	F
Difficulty	0.5563	0.3095	(19, 65)	1.534
Discrimination	0.4393	0.1930	(16, 68)	1.017

*p<0.05

Regression weights for these two models can be found in Table 11. Using these regression weights, item difficulty and item discrimination values were calculated using a similar process as with the LLTM. These estimates can be found in Table C4. The average item difficulty was -0.7547, indicating that the items are, on average, relatively easy. The average item discrimination was 1.3089, indicating that the items did attempt to discrimination between ability levels. These estimated item difficulty and item discrimination values were correlated with their respective values from BILOG; these correlations can be found in Table 9 above.

Table 11. *Regression weights for the 2PL model.*

Parameter	Estimate	Standard Error	t-value
DIFFICULTY			
Math	-0.0009	0.0096	-0.09
Predicate Prop	0.0746	0.0473	1.58
Modifier Prop	-0.0477	0.0410	-1.16
Connective Prop	0.0849	0.1261	0.67
Total Argument Ratio	0.5912	0.3503	1.69
Translate Word Equation	-0.2925	0.2978	-0.98

Table 11 (continued)

Given Equation	-0.3902	0.3613	-1.08
Generate Eq.	-0.0406	0.2688	-0.15
Access Eq.	-0.0138	0.3990	-0.04
Auxiliary Diagram	-0.0042	0.4550	-0.01
Translate Diagram	0.5489	0.3208	1.71
Visualization	0.3302	0.4249	0.78
Semantic Memory	-0.1618	0.1238	-1.31
Number of Subgoals	0.2588	0.1952	1.33
Relative Definition	0.8990	0.6726	1.34
Number Knowledge	0.0297	0.0663	0.45
Procedural Knowledge	0.1183	0.0617	1.92
Number of Operands	0.1569	0.1343	1.17
Decision Processing Confirmation	0.1670	0.2750	0.61
(B Intercept)	-2.0454	0.4511	-4.53*
DISCRIMINATION			
Math	0.0028	0.0040	0.70
LSA	-0.2054	0.1748	-1.18
Translate Word Equation	-0.0050	0.1332	-0.04
Given Equation	0.1503	0.1670	0.90
Generate Eq.	0.1247	0.1294	0.96
Access Eq.	0.1321	0.1912	0.69
Auxiliary Diagram	0.2386	0.2011	1.19
Translate Diagram	-0.1029	0.1461	-0.70
Visualization	0.2322	0.1822	1.28
Semantic Memory	-0.0087	0.0563	-0.16
Number of Subgoals	-0.0850	0.0891	-0.95
Relative Definition	-0.3658	0.3101	-1.18
Number Knowledge	-0.0255	0.0311	-0.82
Procedural Knowledge	0.0217	0.0285	0.76
Number of Operands	-0.0401	0.0617	-0.65
Decision Processing Confirmation	-0.1026	0.1228	-0.84
(A Intercept)	1.4215	0.2162	6.58*

*$p < 0.05$

Similar to the LLTM, examining the regression weights of the cognitive attributes can highlight which attributes contribute to item difficulty and increased burden on working memory. The total argument ratio of the item, need to translate a diagram, visualize a representation, number of subgoals, presence of relative definition, procedural knowledge, the number of operands, and decision processing confirmation led to increased item difficulty and demand placed on working memory. The translating a word equation, given an equation, and the semantic memory requirements of the item led to reduced item difficulty and demand. The reaming attributes, mathematical encoding, the number of predicate propositions, the number of modifier propositions, the number of connective propositions, generate equation, access equation, auxiliary diagram, and number knowledge, had no major impact on item difficulty or demand on working memory.

The influence of the cognitive attributes on item discrimination can also be determined by examining the regression weights. A given equation, generating an equation or possible value, accessing an equation, an auxiliary diagram, and visualizing a representation increased the discrimination of the item. LSA, translating a diagram, presence of relative definitions, and decision processing confirmation decreased the discrimination of the item. The remaining attributes (mathematical encoding, translating a word equation, semantic memory requirements, number of subgoals, number knowledge, procedural knowledge, the number of operands) do not have much impact on discrimination.

4.3 Component Difficulty

The difficulty for each item based on component can also be calculated using the stimulus feature weights from the LLTM and 2PL target models. These component difficulties were estimated in a similar way as total item difficulty except that only the feature weights and attribute scores were used for the attributes within a specific component. For example, to find the difficulty of the translation component, the stimulus features for contextual encoding, number of predicate propositions, number of modifier propositions, number of connective propositions, and total argument ratio were multiplied by their respective scores on items. Then, these values were summed up, including the intercept, to equal the total translation component difficulty for each item. Means and standard deviations for these component difficulties and discrimination can be found in Table 12. For the LLTM, the Translation and Solution Execution components contributed the most to item difficulty whereas the Integration component contributed least to item difficulty. For the 2PL regressions, the Translation and Solution Execution components contributed most to item difficulty whereas the Integration component contributed least to item difficulty. The same components contribute to item difficulty in similar ways in both models.

Table 12. *Means and standard deviations of component difficulties.*

Component	LLTM		2PL Difficulty	
	M	*SD*	*M*	*SD*
Translation	-1.7548	0.3985	-1.3635	0.3755
Integration	-2.7476	0.3965	-2.3866	0.3568
Solution Planning	-2.3413	0.2195	-1.9877	0.1913
Solution Execution	-1.8574	0.3203	-1.2287	0.4211
Decision Processing	-2.2951	0.1184	-1.9609	0.0840

Individual items also vary on difficulty within a component. This can observed by comparing items on two components at once (i.e., comparing the items on Translation difficulty to Integration difficulty). Graphical representations for items on component difficulty in the LLTM and 2PL regressions can be found in Figures D1 through D20. The correlation between the two components is provided in the upper right-hand corner of the plot. The lines represent the mean of the component. As can be seen through the figures, there was more spread in difficulty when comparing Translation, Integration, and Solution Execution component difficulties (i.e., Figure D1 or Figure D16), and less spread when comparing Solution Planning or Decision Processing component difficulties (i.e., Figure D4 or Figure D7).

CHAPTER 5
DISCUSSION

A primary goal of this research is to find a cognitive variable model that assesses the burden imposed on working memory when solving a mathematics assessment item, with an emphasis on contextual variables that best measure the contextualization of the item. A full structure item-response model, the LLTM, is applied to the data to model item difficulty based on cognitive attributes that increase the difficulty of the item and the demand placed on working memory. Unfortunately, the 2PL Constrained was unable to be modeled due to the feasibility of joint constraints on item difficulty and item discrimination (i.e., these parameters are estimated simultaneously). This feasibility should be investigated. Therefore, 2PL regression substitutes were conducted instead. These models measure the burden placed on working memory; as this burden increases, so does the difficulty of the item.

Results of the comparison between the variables show that there are cognitive attributes that significantly contribute to the difficulty of items. Several attributes increased item difficulty in both models. These attributes are: total argument ratio, translating a diagram, number of subgoals, presence of relative definitions, number of operands, and decision processing confirmation. Since the attributes are related to working memory, these attributes contribute to a higher demand placed on working memory. Other attributes are associated with decreased item difficulty, such as: translate word equation, given equation, and semantic memory requirements. These attributes lead to a decreased demand on working memory.

Items can also be compared on their total difficulty or on their difficulty within the five components. As depicted, there is more spread between items within the Translation, Integration, and Solution Execution components. Spread drastically drops when looking at the Solution Planning and Decision Processing components. As seen in Figure D9, it appears there are only eight items on the exam. However, there are 85 items but the items are so similar on these two components there is no spread within the figure. Or, as with Figure D14, the only source of item difficulty variation is within translation, not within decision processing. The Translation and Solution Execution components lead to more spread in item difficulty, increasing difficulty within the items. The Integration component also has a lot of spread but decreases the difficulty of the items.

An area of interest within the creation of this model was modeling the extent of contextualization of the mathematics items. A traditional propositional analysis was used in the analysis because it focuses more on the underlying structure of the item rather than the surface structure. Also, this analysis is more adaptable when examining mathematical items over other analyses, such as LSA and CPIDR 3. These later analyses, although ideal because they are automated, do not seem to be geared towards mathematical assessment. A prime example of this is the misclassification of the noun “set” as a verb in CPIDR 3. LSA is more related to longer texts, suggesting it is a more global variable.

The mathematics items on the current exam used for this study do not seem to be heavily contextualized. Contextual variables do contribute to item difficulty but have smaller influences than predicted. This could be because the current items are not heavily contextualized now, but with current directions, this could become a problem soon. The Translation component does show a wide range of difficulty without these contextual variables having large weights. If the trend continues and mathematics items are more highly contextualized, this difficulty is going to explode, which may increase difficulty and will certainly lead to an even wider range. Having a model in place that could measure this impact would be extremely beneficial.

Contextualization of mathematical items is, as previously stated, becoming more common. By creating mathematical items based on real-world scenarios (i.e., contextualization), item writers may be steering away from strictly measuring math ability. In highly contextualized items, other abilities, such as reading ability and comprehension, may have more influential roles in the difficulty of the item and the process by which a solution is found. Instead of measuring math ability, these more contextualized items may be measuring something else, such as general intelligence (*Gf*). This notion is also supported by the fact that the variables within these models are related to working memory, and working memory has been shown to be related to *Gf* (Engle et al., 1999; Conway et al., 2002; Ackerman et al., 2005; Unsworth & Engle, 2007). Even mathematics performance itself has been shown to relate to working memory and *Gf* (Floyd, Evans, McGrew, 2003).

With this shift in item construction, item difficulty may not be the only issue affected. The concept now being measured may also be altered. Therefore, an end-of-the-year mathematics assessment exam may no longer measure a child’s mastery of mathematics; the results of the exam may be confounded with a child’s mastery of other skills. For example, an item that places mathematical constructs within a word problem may be measuring a child’s reading ability (i.e., can he or she read and understand the story, getting rid of irrelevant information) rather than his or her mathematical ability (i.e., can he or she add together the appropriate numbers in the correct order).

This cognitive model is also beneficial in several ways. First, it provides a moderate prediction of the difficulty and discrimination of a mathematics item by assessing sources of burden imposed on working memory by the item. Second, the extent of contextualization of the item is observed. Although contextualization is moderate in the items used for this analysis, if programs such as CBAL and those proposed by SBAC are implemented, contextualization could become a greater problem. Having a model that can report on the amount of contextualization of an item would be beneficial to item developers, teachers, parents, and students. Third, the model will help inform test design procedures for computer-based, computer-adaptive, and standard testing. Item writers can use the model to create items of similar difficulty to the original item or of varying difficulty (higher or lower) to the original item. Educational implications, such as these, will only continue to grow in magnitude. Models such as the one in this paper will be of benefit to students, teachers, parents, states, and the nation, which is a major goal of educational programs, such as PARCC, being created, evaluated, and implemented today.

APPENDIX A
IRT ESTIMATION

Table A 1. *Comparison of IRT models.*

Model	K	AIC	-2 Log-Likelihood	Δ -2 Log-Likelihood	Additional Parameters
Rasch	85	237,602.6394	237,432.6394		
1PL	85	237,602.6394	237,432.6394	0	0
2PL	170	235,203.3768	234,863.3768	2,569.2626	85
3PL Common	171	234,365.4628	234,023.4628	839.914	1
3PL	255	234,106.0794	233,596.0794	427.3834	84

Table A 2. *Classical item analysis.*

Item	Attempted	Correct	Percent Correct	Pearson	Biserial
1	2993	2867	95.8	0.209	0.465
2	2993	2615	87.4	0.302	0.484
3	2993	1704	56.9	0.284	0.358
5	2993	2423	81	0.486	0.702
6	2993	1869	62.4	0.499	0.637
7	2993	2825	94.4	0.212	0.43
8	2993	2628	87.8	0.383	0.62
9	2993	1689	56.4	0.517	0.651
10	2993	1715	57.3	0.353	0.445
11	2993	2507	83.8	0.329	0.493
12	2993	1163	38.9	0.293	0.372
13	2993	1686	56.3	0.366	0.461
14	2993	2780	92.9	0.258	0.487
15	2993	2429	81.2	0.417	0.604
16	2993	2019	67.5	0.387	0.503
17	2993	2406	80.4	0.397	0.57
18	2993	2525	84.4	0.445	0.674
19	2993	1910	63.8	0.494	0.633
20	2993	2155	72	0.403	0.538
21	2993	2129	71.1	0.561	0.744
22	2993	2070	69.2	0.542	0.711
23	2993	2347	78.4	0.473	0.665
24	2993	2431	81.2	0.387	0.562
25	2993	1999	66.8	0.479	0.621
26	2993	1304	43.6	0.443	0.558
27	2993	2257	75.4	0.367	0.502
28	2993	1729	57.8	0.431	0.543
29	2993	2249	75.1	0.495	0.676
30	2993	1931	64.5	0.499	0.641
31	2993	2629	87.8	0.298	0.483
32	2993	2293	76.6	0.408	0.564
33	2993	1740	58.1	0.357	0.451
34	2993	2786	93.1	0.315	0.601
35	2993	2261	75.5	0.493	0.674
36	2993	2287	76.4	0.262	0.362
37	2993	2773	92.6	0.367	0.686
38	2993	1618	54.1	0.242	0.303
39	2993	2553	85.3	0.441	0.679
40	2993	2497	83.4	0.453	0.676
41	2993	1952	65.2	0.442	0.569

Item	Attempted	Correct	Percent Correct	Pearson	Biserial
42	2993	2578	86.1	0.431	0.674
43	2993	2530	84.5	0.35	0.531
44	2993	1904	63.6	0.533	0.682
45	2993	1886	63	0.532	0.68
46	2993	2367	79.1	0.449	0.635
47	2993	1711	57.2	0.414	0.522
48	2993	1493	49.9	0.425	0.532
49	2993	2507	83.8	0.417	0.625
50	2993	1006	33.6	0.504	0.652
51	2993	2828	94.5	0.36	0.736
52	2993	1787	59.7	0.572	0.725
53	2993	1560	52.1	0.51	0.64
54	2993	987	33	0.298	0.387
55	2993	1354	45.2	0.414	0.52
56	2993	1633	54.6	0.404	0.507
57	2993	1235	41.3	0.392	0.496
58	2993	2796	93.4	0.289	0.56
59	2993	2692	89.9	0.261	0.445
60	2993	2651	88.6	0.334	0.551
61	2993	2171	72.5	0.374	0.501
62	2993	2501	83.6	0.277	0.414
63	2993	1034	34.5	0.409	0.528
64	2993	2498	83.5	0.452	0.675
65	2993	2769	92.5	0.234	0.435
66	2993	2258	75.4	0.349	0.477
67	2993	1688	56.4	0.403	0.508
68	2993	2009	67.1	0.399	0.518
69	2993	2078	69.4	0.475	0.625
70	2993	1940	64.8	0.446	0.574
71	2993	2033	67.9	0.459	0.598
72	2993	1943	64.9	0.435	0.56
73	2993	2306	77	0.28	0.388
74	2993	2056	68.7	0.289	0.378
75	2993	2742	91.6	0.355	0.639
76	2993	2315	77.3	0.374	0.521
77	2993	2635	88	0.367	0.597
78	2993	1835	61.3	0.447	0.569
79	2993	1912	63.9	0.47	0.603
80	2993	2535	84.7	0.38	0.58
81	2993	2116	70.7	0.431	0.57
82	2993	2130	71.2	0.417	0.553
83	2993	2445	81.7	0.51	0.743

Item	Attempted	Correct	Percent Correct	Pearson	Biserial
84	2993	1490	49.8	0.383	0.48
85	2993	2200	73.5	0.364	0.491
86	2993	862	28.8	0.173	0.229

Table A 3. *Parameter estimates and their standard errors for 1PL and 2PL models.*

	1PL				2PL			
Item	*a*	*s{a}*	*b*	*s{b}*	*a*	*s{a}*	*b*	*s{b}*
p1i1	1.181	0.006	-3.123	0.08	1.263	0.116	-2.964	0.209
p1i2	1.181	0.006	-2.026	0.049	1.132	0.08	-2.083	0.112
p1i3	1.181	0.006	-0.325	0.033	0.67	0.044	-0.478	0.062
p1i5	1.181	0.006	-1.549	0.046	1.848	0.087	-1.228	0.039
p1i6	1.181	0.006	-0.573	0.037	1.428	0.061	-0.536	0.032
p1i7	1.181	0.006	-2.852	0.069	1.12	0.109	-2.954	0.225
p1i8	1.181	0.006	-2.064	0.052	1.609	0.09	-1.714	0.064
p1i9	1.181	0.006	-0.303	0.037	1.516	0.063	-0.291	0.029
p1i10	1.181	0.006	-0.342	0.034	0.859	0.047	-0.423	0.048
p1i11	1.181	0.006	-1.741	0.045	1.077	0.064	-1.852	0.092
p1i12	1.181	0.006	0.479	0.034	0.725	0.043	0.684	0.067
p1i13	1.181	0.006	-0.299	0.034	0.907	0.048	-0.358	0.045
p1i14	1.181	0.006	-2.621	0.063	1.246	0.099	-2.517	0.149
p1i15	1.181	0.006	-1.562	0.044	1.403	0.068	-1.414	0.056
p1i16	1.181	0.006	-0.809	0.036	1.001	0.051	-0.906	0.053
p1i17	1.181	0.006	-1.512	0.043	1.295	0.07	-1.434	0.061
p1i18	1.181	0.006	-1.785	0.048	1.77	0.088	-1.428	0.047
p1i19	1.181	0.006	-0.637	0.038	1.468	0.064	-0.585	0.032
p1i20	1.181	0.006	-1.035	0.038	1.184	0.06	-1.043	0.049
p1i21	1.181	0.006	-0.991	0.041	2.006	0.09	-0.789	0.027
p1i22	1.181	0.006	-0.892	0.04	1.825	0.078	-0.739	0.028
p1i23	1.181	0.006	-1.391	0.043	1.71	0.083	-1.148	0.039
p1i24	1.181	0.006	-1.566	0.044	1.305	0.073	-1.477	0.062
p1i25	1.181	0.006	-0.777	0.038	1.362	0.06	-0.732	0.037
p1i26	1.181	0.006	0.264	0.036	1.153	0.052	0.254	0.038
p1i27	1.181	0.006	-1.218	0.039	1.065	0.06	-1.307	0.064
p1i28	1.181	0.006	-0.362	0.035	1.133	0.051	-0.384	0.038
p1i29	1.181	0.006	-1.203	0.042	1.652	0.076	-1.016	0.036
p1i30	1.181	0.006	-0.669	0.038	1.474	0.063	-0.613	0.032
p2i1	1.181	0.006	-2.067	0.05	1.066	0.067	-2.216	0.115
p2i2	1.181	0.006	-1.285	0.041	1.211	0.061	-1.272	0.057
p2i3	1.181	0.006	-0.379	0.034	0.871	0.048	-0.465	0.049
p2i4	1.181	0.006	-2.649	0.066	1.691	0.106	-2.108	0.087
p2i5	1.181	0.006	-1.225	0.042	1.662	0.077	-1.032	0.037
p2i6	1.181	0.006	-1.274	0.038	0.681	0.05	-1.906	0.134
p2i7	1.181	0.006	-2.589	0.065	2.22	0.147	-1.812	0.061
p2i8	1.181	0.006	-0.198	0.032	0.577	0.042	-0.32	0.068

	1PL				2PL			
Item	*a*	*s{a}*	*b*	*s{b}*	*a*	*s{a}*	*b*	*s{b}*
p2i9	1.181	0.006	-1.856	0.05	1.861	0.091	-1.448	0.046
p2i10	1.181	0.006	-1.717	0.047	1.804	0.091	-1.365	0.044
p2i11	1.181	0.006	-0.702	0.037	1.229	0.059	-0.699	0.038
p2i12	1.181	0.006	-1.922	0.05	1.833	0.094	-1.505	0.048
p2i13	1.181	0.006	-1.797	0.046	1.237	0.075	-1.745	0.081
p2i14	1.181	0.006	-0.627	0.038	1.675	0.072	-0.549	0.028
p2i15	1.181	0.006	-0.6	0.038	1.627	0.068	-0.532	0.029
p2i16	1.181	0.006	-1.431	0.043	1.566	0.08	-1.229	0.044
p2i17	1.181	0.006	-0.336	0.035	1.056	0.05	-0.37	0.04
p2i18	1.181	0.006	-0.015	0.035	1.115	0.052	-0.028	0.038
p2i19	1.181	0.006	-1.741	0.047	1.604	0.091	-1.462	0.053
p2i20	1.181	0.006	0.728	0.04	1.607	0.062	0.589	0.035
p2i21	1.181	0.006	-2.869	0.074	2.748	0.204	-1.824	0.056
p2i22	1.181	0.006	-0.449	0.039	1.785	0.072	-0.397	0.027
p2i23	1.181	0.006	-0.113	0.037	1.428	0.059	-0.127	0.031
p2i24	1.181	0.006	0.759	0.036	0.83	0.044	0.967	0.071
p2i25	1.181	0.006	0.19	0.035	1.088	0.052	0.188	0.04
p2i26	1.181	0.006	-0.22	0.035	1.061	0.054	-0.246	0.039
p2i27	1.181	0.006	0.368	0.036	1.068	0.05	0.382	0.045
p3i1	1.181	0.006	-2.698	0.067	1.508	0.102	-2.29	0.11
p3i2	1.181	0.006	-2.27	0.054	1.017	0.075	-2.518	0.152
p3i3	1.181	0.006	-2.135	0.053	1.313	0.077	-1.99	0.09
p3i4	1.181	0.006	-1.063	0.038	1.045	0.057	-1.156	0.059
p3i5	1.181	0.006	-1.726	0.044	0.852	0.059	-2.179	0.134
p3i6	1.181	0.006	0.682	0.037	1.133	0.051	0.687	0.046
p3i7	1.181	0.006	-1.719	0.047	1.783	0.089	-1.374	0.044
p3i8	1.181	0.006	-2.571	0.062	1.006	0.077	-2.882	0.187
p3i9	1.181	0.006	-1.22	0.039	0.96	0.055	-1.403	0.075
p3i10	1.181	0.006	-0.302	0.035	1.015	0.05	-0.34	0.041
p3i11	1.181	0.006	-0.793	0.036	1.081	0.056	-0.847	0.047
p3i12	1.181	0.006	-0.905	0.039	1.423	0.067	-0.83	0.037
p3i13	1.181	0.006	-0.683	0.037	1.208	0.057	-0.687	0.04
p3i14	1.181	0.006	-0.831	0.038	1.343	0.067	-0.787	0.037
p3i15	1.181	0.006	-0.688	0.037	1.177	0.058	-0.702	0.04
p3i16	1.181	0.006	-1.31	0.038	0.794	0.057	-1.736	0.112
p3i17	1.181	0.006	-0.869	0.035	0.708	0.047	-1.25	0.089
p3i18	1.181	0.006	-2.457	0.061	1.846	0.119	-1.879	0.071
p3i19	1.181	0.006	-1.328	0.04	1.124	0.063	-1.375	0.066
p3i20	1.181	0.006	-2.085	0.052	1.53	0.093	-1.778	0.072

Item	1PL				2PL			
	a	*s{a}*	*b*	*s{b}*	*a*	*s{a}*	*b*	*s{b}*
p3i21	1.181	0.006	-0.522	0.036	1.217	0.058	-0.527	0.037
p3i22	1.181	0.006	-0.64	0.037	1.327	0.06	-0.615	0.035
p3i23	1.181	0.006	-1.81	0.047	1.399	0.078	-1.632	0.066
p3i24	1.181	0.006	-0.969	0.038	1.241	0.062	-0.952	0.045
p3i25	1.181	0.006	-0.993	0.038	1.194	0.06	-0.995	0.048
p3i26	1.181	0.006	-1.597	0.047	2.131	0.106	-1.196	0.033
p3i27	1.181	0.006	-0.011	0.034	0.964	0.047	-0.018	0.042
p3i28	1.181	0.006	-1.114	0.038	1.024	0.058	-1.228	0.063
p3i29	1.181	0.006	0.972	0.035	0.514	0.039	1.861	0.158

Table A 4. *Parameter estimates and their standard errors for 3PLC and 3PL models.*

	3PL Common Estimates (c = 0.188, $s\{c\}$ = .002)				3PL Estimates					
Item	a	$s\{a\}$	b	$s\{b\}$	a	$s\{a\}$	b	$s\{b\}$	c	$s\{c\}$
p1i1	1.139	0.101	-3.036	0.211	1.116	0.1	-3.032	0.265	0.235	0.092
p1i2	1.056	0.077	-1.919	0.114	1.413	0.151	-0.999	0.218	0.5	0.066
p1i3	0.857	0.06	0.182	0.057	1.166	0.15	0.587	0.129	0.313	0.039
p1i5	1.879	0.1	-0.991	0.045	1.812	0.118	-1.028	0.088	0.158	0.047
p1i6	1.916	0.1	-0.112	0.032	2.02	0.139	-0.055	0.057	0.208	0.026
p1i7	0.996	0.096	-3.015	0.234	0.973	0.097	-2.955	0.315	0.265	0.1
p1i8	1.521	0.09	-1.574	0.07	1.511	0.108	-1.539	0.146	0.217	0.072
p1i9	2.493	0.145	0.14	0.027	2.765	0.208	0.206	0.038	0.219	0.019
p1i10	1.138	0.069	0.136	0.046	1.379	0.136	0.356	0.096	0.268	0.033
p1i11	1.067	0.066	-1.578	0.09	1.068	0.081	-1.53	0.204	0.212	0.077
p1i12	1.434	0.106	1.094	0.055	1.649	0.182	1.117	0.06	0.214	0.019
p1i13	1.244	0.074	0.185	0.043	1.551	0.146	0.403	0.08	0.273	0.029
p1i14	1.111	0.09	-2.527	0.159	1.107	0.096	-2.424	0.255	0.268	0.098
p1i15	1.412	0.075	-1.157	0.059	1.36	0.081	-1.258	0.108	0.126	0.049
p1i16	1.144	0.064	-0.441	0.05	1.146	0.092	-0.445	0.14	0.18	0.051
p1i17	1.312	0.077	-1.147	0.062	1.311	0.101	-1.11	0.153	0.204	0.065
p1i18	1.732	0.094	-1.233	0.054	1.693	0.112	-1.254	0.107	0.174	0.055
p1i19	1.975	0.105	-0.158	0.032	2.233	0.161	-0.034	0.054	0.245	0.026
p1i20	1.321	0.073	-0.636	0.048	1.504	0.127	-0.391	0.115	0.284	0.046
p1i21	2.371	0.131	-0.458	0.03	2.366	0.168	-0.444	0.055	0.179	0.031
p1i22	2.307	0.122	-0.373	0.03	2.414	0.162	-0.312	0.051	0.21	0.027
p1i23	1.831	0.1	-0.851	0.042	1.995	0.155	-0.689	0.088	0.268	0.043
p1i24	1.33	0.079	-1.185	0.062	1.56	0.144	-0.791	0.148	0.355	0.058
p1i25	1.596	0.082	-0.34	0.039	1.525	0.101	-0.418	0.082	0.137	0.036
p1i26	1.817	0.111	0.666	0.036	1.609	0.118	0.526	0.051	0.113	0.02

Item	3PL Common Estimates (c = 0.188, $s\{c\}$ = .002)				3PL Estimates					
	a	$s\{a\}$	b	$s\{b\}$	a	$s\{a\}$	b	$s\{b\}$	c	$s\{c\}$
p1i27	1.146	0.069	-0.904	0.06	1.335	0.13	-0.541	0.159	0.322	0.056
p1i28	1.601	0.086	0.097	0.036	1.779	0.132	0.197	0.063	0.23	0.026
p1i29	1.83	0.096	-0.689	0.039	1.9	0.135	-0.611	0.081	0.221	0.04
p1i30	1.96	0.102	-0.192	0.032	2.139	0.148	-0.095	0.056	0.23	0.026
p2i1	1.038	0.066	-2	0.113	1.038	0.074	-1.958	0.213	0.213	0.082
p2i2	1.267	0.069	-0.939	0.058	1.217	0.076	-1.061	0.114	0.12	0.047
p2i3	1.105	0.069	0.079	0.046	1.155	0.117	0.133	0.126	0.204	0.044
p2i4	1.502	0.097	-2.122	0.098	1.474	0.103	-2.111	0.166	0.224	0.085
p2i5	1.826	0.097	-0.712	0.04	1.858	0.134	-0.668	0.084	0.202	0.042
p2i6	0.73	0.054	-1.362	0.111	0.732	0.065	-1.341	0.277	0.193	0.075
p2i7	1.934	0.135	-1.821	0.073	1.951	0.173	-1.704	0.151	0.297	0.084
p2i8	0.767	0.059	0.418	0.067	1.562	0.219	0.998	0.084	0.383	0.024
p2i9	1.804	0.097	-1.271	0.052	1.773	0.117	-1.264	0.108	0.194	0.058
p2i10	1.836	0.102	-1.131	0.048	1.993	0.155	-0.946	0.103	0.29	0.051
p2i11	1.572	0.086	-0.238	0.037	1.951	0.158	0.005	0.069	0.293	0.029
p2i12	1.772	0.098	-1.331	0.054	1.874	0.139	-1.163	0.118	0.287	0.06
p2i13	1.21	0.077	-1.513	0.081	1.219	0.098	-1.431	0.19	0.231	0.078
p2i14	2.446	0.14	-0.131	0.028	2.797	0.216	-0.019	0.043	0.244	0.022
p2i15	2.19	0.115	-0.133	0.03	2.271	0.152	-0.098	0.049	0.194	0.025
p2i16	1.656	0.093	-0.926	0.046	1.967	0.166	-0.623	0.097	0.33	0.044
p2i17	1.422	0.079	0.121	0.039	1.513	0.122	0.181	0.077	0.209	0.03
p2i18	1.834	0.109	0.434	0.034	2.05	0.168	0.495	0.049	0.217	0.02
p2i19	1.594	0.097	-1.23	0.056	2.029	0.189	-0.754	0.11	0.411	0.047
p2i20	3.37	0.258	0.857	0.027	2.711	0.181	0.72	0.028	0.08	0.01
p2i21	2.291	0.177	-1.903	0.071	2.328	0.23	-1.774	0.143	0.329	0.085
p2i22	2.296	0.121	-0.036	0.029	2.076	0.121	-0.151	0.044	0.095	0.022

Item	3PL Common Estimates (c = 0.188, $s\{c\}$ = .002)				3PL Estimates					
	a	$s\{a\}$	b	$s\{b\}$	a	$s\{a\}$	b	$s\{b\}$	c	$s\{c\}$
p2i23	2.034	0.114	0.275	0.031	1.883	0.122	0.175	0.046	0.12	0.02
p2i24	3.032	0.273	1.129	0.034	3.478	0.37	1.11	0.033	0.2	0.011
p2i25	2.105	0.143	0.625	0.033	2.727	0.27	0.71	0.038	0.237	0.016
p2i26	1.693	0.104	0.267	0.035	2.847	0.281	0.552	0.042	0.323	0.018
p2i27	2.855	0.215	0.761	0.029	3.586	0.349	0.804	0.031	0.226	0.013
p3i1	1.365	0.094	-2.293	0.119	1.313	0.095	-2.349	0.178	0.195	0.079
p3i2	0.98	0.072	-2.336	0.147	0.973	0.076	-2.306	0.24	0.215	0.085
p3i3	1.264	0.076	-1.831	0.093	1.23	0.079	-1.907	0.148	0.154	0.063
p3i4	1.164	0.069	-0.719	0.054	1.378	0.13	-0.368	0.139	0.316	0.049
p3i5	0.87	0.061	-1.811	0.12	0.865	0.069	-1.781	0.249	0.205	0.08
p3i6	2.384	0.179	0.993	0.035	2.067	0.166	0.88	0.04	0.123	0.014
p3i7	1.777	0.097	-1.155	0.049	1.853	0.137	-1.038	0.109	0.252	0.055
p3i8	0.98	0.073	-2.712	0.176	0.968	0.074	-2.724	0.242	0.2	0.081
p3i9	1.018	0.062	-0.995	0.07	0.994	0.073	-1.068	0.167	0.155	0.059
p3i10	1.402	0.081	0.164	0.039	1.595	0.138	0.285	0.074	0.236	0.029
p3i11	1.286	0.073	-0.38	0.044	1.527	0.138	-0.113	0.103	0.289	0.039
p3i12	1.712	0.094	-0.426	0.037	1.98	0.161	-0.235	0.073	0.273	0.033
p3i13	1.417	0.076	-0.265	0.041	1.344	0.094	-0.366	0.093	0.132	0.039
p3i14	1.642	0.094	-0.362	0.037	2.127	0.191	-0.068	0.068	0.315	0.03
p3i15	1.461	0.081	-0.244	0.039	1.72	0.146	-0.039	0.081	0.27	0.033
p3i16	0.826	0.061	-1.279	0.097	1.117	0.147	-0.371	0.237	0.442	0.061
p3i17	0.799	0.056	-0.659	0.073	0.846	0.092	-0.483	0.256	0.236	0.072
p3i18	1.648	0.11	-1.851	0.081	1.671	0.141	-1.712	0.172	0.296	0.088
p3i19	1.179	0.071	-1.019	0.063	1.186	0.098	-0.975	0.169	0.204	0.066
p3i20	1.439	0.092	-1.636	0.076	1.563	0.142	-1.332	0.181	0.347	0.078
p3i21	1.583	0.09	-0.07	0.036	1.749	0.147	0.04	0.072	0.231	0.031

	3PL Common Estimates (c = 0.188, $s\{c\}$ = .002)				3PL Estimates					
Item	a	$s\{a\}$	b	$s\{b\}$	a	$s\{a\}$	b	$s\{b\}$	c	$s\{c\}$
p3i22	1.732	0.093	-0.174	0.035	1.957	0.149	-0.036	0.065	0.248	0.029
p3i23	1.366	0.08	-1.419	0.07	1.365	0.103	-1.358	0.162	0.224	0.072
p3i24	1.422	0.079	-0.545	0.044	1.504	0.124	-0.434	0.108	0.23	0.045
p3i25	1.351	0.075	-0.586	0.046	1.421	0.117	-0.478	0.118	0.229	0.048
p3i26	2.272	0.131	-0.943	0.038	2.49	0.199	-0.791	0.073	0.276	0.04
p3i27	1.534	0.091	0.48	0.039	1.788	0.154	0.577	0.058	0.234	0.022
p3i28	1.125	0.068	-0.793	0.058	1.412	0.139	-0.32	0.14	0.357	0.048
p3i29	2.204	0.236	1.573	0.061	2.621	0.369	1.547	0.056	0.211	0.011

APPENDIX B
COGNITIVE MODEL VARIABLE DEFINITIONS AND SCORES

Table B 1. *Cognitive variables and their definitions.*

Attribute	Definition	Scale
Translation		
Encoding Total	Total number of words and math terms, both explicit and implicit, in the stem and all answer options. Equal to the sum total of contextual and mathematical attributes.	Ratio : count
Mathematical	Total number of mathematical terms, explicit and implicit, in the stem and all answer options. This includes numerals, variables (e.g., a, y, x, m, etc.), axis labels, comparators (e.g., <, >, =), and implicit and explicit operators.	Ratio : count
Contextual	Total numbers of words, excluding variables, in both the stem and answer options.	Ratio : count
Stem	Total number of words and math terms, both explicit and implicit, in the stem.	Ratio : count
Content Words	The number of content words (excluding articles such as the, it, an, etc.) within the stem of an item – these are words that serve more than just a linguistic purpose	Ratio : count
Text Comprehension		
Flesch Reading Ease Test	The readability of the problem stem – the ease of reading.	Interval
Flesch-Kincaid Grade Level Test	The maximum MS-word reading level for the stem of a problem.	Interval
Text Comprehension – LSA	Comparison of the similarity of sequential sentences within the LSA space - each sentence is compared to the next The LSA space is general reading up to the 9th grade.	Interval

Attribute		Definition	Scale
		This is the mean of all comparisons between sequential sentences.	
	Mathematical Propositions		
	Assignment Propositions	Indicator of whether a numeric value has been assigned to a variable	Ratio : binary
	Relation Propositions	Indicator of whether a numeric relation is expressed between two variables	Ratio : binary
	Contextual Propositions		
	Total Number of Propositions	The total number of predicate, connector, and modifier propositions that occur within the stem of an item	Ratio : count
	Total Proposition Density	Total number of propositions divided by the word count within the stem	Interval
	Number of Predicate Propositions	Number of unique Predicate Propositions in the stem of an item	Ratio : count
	Predicate Density	Number of unique Predicate propositions within the stem divided by the word count within the stem	Interval
	Number of Modifier Propositions	Number of unique Modifier Propositions in the stem of an item	Ratio : count
	Modifier Density	Number of unique Modifier propositions within the stem divided by the word count within the stem	Interval
	Number of Connector Propositions	Number of unique Connector Propositions in the stem of an item	Ratio : count
	Connector Density	Number of unique Connector propositions within the stem divided by the word count within the stem	Interval
	Total Number of Unique Arguments	Total number of unique arguments that appear within	Ratio : count

Attribute		Definition	Scale
		the stem of an item	
	Unique Argument Density	The total number of unique arguments within an item's stem divided by the word count within the stem	Interval
	Total Number of Arguments	Total number of arguments that occur within the stem of an item	Ratio : count
	Total Argument Ratio	The total number of arguments within an item's stem divided by the word count within the stem	Interval
	Max. Number of Arguments	Maximum number of arguments within a proposition in the stem of an item	Ratio : count
	Relevant Propositions	The number of mathematically relevant propositions within the stem of an item	Ratio : count
	Density of Relevant Propositions	Total number of relevant propositions within the stem of an item divided by the total number of propositions in an item	Interval
	Relevant Words	Total number of relevant words within the stem of an item	Ratio : count
	Density of Relevant Words	Total number of relevant words within the stem of an item divided by the total number of words within an item	Interval
	Irrelevant Propositions	Total number of mathematically irrelevant propositions within the stem of an item	Ratio : count
	Density of Irrelevant Propositions	Total number of irrelevant propositions within the stem of an item divided by the total number of propositions within an item	Interval
	Irrelevant Words	Total number of irrelevant words within the stem of an item	Ratio : count
	Density of Irrelevant Words	Total number of irrelevant words within the stem of an item divided by the total number of words within an item	Interval
	Encode Diagram	Indicator of presence of a	Ratio :

Attribute		Definition	Scale
		diagram, graph, or other figure, excluding tables, in the stem or answer options.	binary
Integration			
	Translate Word Equation	Indicator of whether the examinee needs to interpret an equation given in word (context) form.	Ratio : binary
	Given Equation – in Stem	Indicator of whether a mathematical equation is given in the stem of a problem.	Ratio : binary
	Generate Eq. or Possible Values	Indicator of whether examinee must generate or derive equations or possible values in order to answer the question. This includes the translation of linguist information to a mathematical format.	Ratio : binary
	Access Equation	Indicator of whether the examinee must access the equation (e.g., Pythagorean theorem, area of a circle, etc.) from a drop-down box.	Ratio : binary
	Auxiliary Diagram	Indicator for whether the presented diagram is unnecessary for problem solution.	Ratio : binary
	Translate Diagram	Indicator for whether presented diagram or figure is necessary for problem solution.	Ratio : binary
	Visualization	Indicator of whether an examinee must draw or otherwise visualize a representation in order to understand or answer a problem.	Ratio : binary
	Semantic Memory	The total number of unique indicators used within an item	Ratio : count
Solution Planning			
	Presence of Subgoals	Indicator of whether subgoals are necessary for solving a problem.	Ratio : binary
	Number of Subgoals	The total number of sub-steps	Ratio : count

Attribute		Definition	Scale
		needed for solving a problem.	
	Relative Definition of Variables	Indicator of whether one variable is defined only in terms of another.	Ratio : binary
Solution Execution			
	Number Knowledge	The maximum number knowledge required for solving an item.	Ordinal
	1. Single-digit	Indicator of whether the problem involves the use single-digit numbers.	Ratio : binary
	2. Double-digit	Indicator of whether the problem involves the use of double-digit numbers.	Ratio : binary
	3. Triple-digit	Indicator of whether the problem involves the use of triple-digit numbers.	Ratio : binary
	4. Four-digit +	Indicator of whether the problem involves the use of four-digit or more numbers.	Ratio : binary
	5. Fraction/Decimal	Indicator of whether the problem involves the use of fractions or decimals.	Ratio : binary
	Alt. Procedural Knowledge	The maximum procedural knowledge needed in solving an item.	Ordinal
	1. Multiple Steps	Does the item require multiple steps in order to solve the problem	Ratio : binary
	2. Algebraic Equations	Does the equation have algebraic equations included in the stem/answer option	Ratio : binary
	3. Mixed Fractions	Does the item require the use of mixed fractions to solve the item	Ratio : binary
	Procedural Knowledge	The maximum procedural knowledge necessary in solving the item	Ordinal
	1. Integers	Indicator for whether ability of integers is necessary for solving	Ratio: binary

Attribute		Definition	Scale
		the item	
	2. Fractions	Indicator for whether ability to manipulate fractions is necessary for solving the item	Ratio: binary
	3. Proportions	Indicator for whether ability to manipulate proportions is necessary for solving the item	Ratio: binary
	4. Decimals	Indicator for whether ability to manipulate decimals is necessary for solving the item	Ratio: binary
	5. Negative Numbers	Indicator for whether ability to manipulate negative numbers is necessary for solving the item	Ratio: binary
	6. Square Roots	Indicator for whether ability to evaluate squares or square roots is necessary for solving the item	Ratio: binary
	Number of Procedures	The total number of procedures necessary for solving the item	Ratio : count
	Number of Computations	The total number of computations necessary for solving an item.	Ratio : count
	Number of Operands	The total number of operands (i.e., adding, subtracting, multiplying, dividing) necessary for solving an item.	Ratio : count
	Meta-Cognition Process	Indicator of whether examinee needs to use metacognition process (i.e., goal setting, self-questioning, self-instruction) to solve the problem.	Ratio : binary
Decision Processing			
	Decision Processing Confirmation	Indicator of whether information found in the distractors is necessary to eliminate options or answer the item.	Ratio : binary
	Bottom-Up Processing	Indicator of whether distractors must be consulted to answer the problem.	Ratio : binary
	Top-Down Processing	Indicator of whether a solution identified from the stem must be compared against information in the distractors to identify the correct answer.	Ratio : binary

Attribute		Definition	Scale
	Functional Distractors	The total number of distractors that was chosen by 5% or more of the examinees.	Ratio : count

Table B 2. *Descriptive statistics for all cognitive variables.*

	Mean	SD	Min	Max
Translation				
Total Encoding	55.0471	27.8627	12.0000	171.0000
Mathematical	19.9412	15.2880	0.0000	97.0000
Contextual	35.1059	20.3873	5.0000	116.0000
Stem	36.7176	17.3827	8.0000	89.0000
Content Words	27.0118	13.7966	5.0000	73.0000
Text Comprehensions				
Flesch Reading Ease Test	74.1694	14.3848	42.6000	100.0000
Flesch-Kincaid Grade Level Test	5.9788	2.5031	0.0000	11.1000
Text Comprehension– LSA (Average)	0.5447	0.3095	-0.0500	1.0000
Comparison One	0.5639	0.3182	-0.0500	1.0000
Comparison Two	0.3686	0.2533	0.0100	0.9200
Comparison Three	0.3715	0.2350	0.0700	0.7400
Comparison Four	0.3933	0.2627	0.1900	0.6900
Mathematical Propositions				
Assignment Propositions	0.2118	0.4110	0.0000	1.0000
Relation Propositions	0.0471	0.2130	0.0000	1.0000
Contextual Propositions				
Total Number of Propositions	10.4588	6.4449	1.0000	30.0000
Total Proposition Density	0.2802	0.1008	0.0652	0.5000
Number of Predicate Propositions	4.1059	2.9601	1.0000	14.0000
Predicate Density	0.1121	0.0491	0.0238	0.2034
Number of Modifier Propositions	5.5529	3.7844	0.0000	17.0000
Modifier Density	0.1462	0.0760	0.0000	0.3333
Number of Connective Propositions	0.8000	0.9735	0.0000	4.0000
Connector Density	0.0219	0.0286	0.0000	0.1000
Total Number of Unique Arguments	10.0235	5.0615	2.0000	25.0000
Unique Argument Density	0.2837	0.0863	0.0952	0.5000
Total Number of Arguments	39.0588	29.9652	2.0000	146.0000
Total Argument Ratio	0.9989	0.4821	0.1522	2.5833
Max. Number of Arguments	7.0118	3.2532	2.0000	18.0000
Relevant Propositions	9.2000	5.4002	1.0000	25.0000
Density of Relevant Propositions	0.9299	0.1771	0.0588	1.0000
Relevant Words	33.8118	14.7846	8.0000	61.0000
Density of Relevant Words	0.9484	0.1334	0.2830	1.0000
Irrelevant Propositions	1.2588	3.7070	0.0000	19.0000
Density of Irrelevant Propositions	0.0701	0.1771	0.0000	0.9412
Irrelevant Words	2.9059	8.3646	0.0000	42.0000

	Mean	SD	Min	Max
Density of Irrelevant Words	0.0516	0.1334	0.0000	0.7170
Encode Diagram	0.2353	0.4267	0.0000	1.0000
Integration				
Translate Word Equation	0.3176	0.4683	0.0000	1.0000
Given Equation – in Stem	0.2353	0.4267	0.0000	1.0000
Generate Eq. or Possible Values	0.4118	0.4951	0.0000	1.0000
Access Equation	0.1412	0.3503	0.0000	1.0000
Auxiliary Diagram	0.0706	0.2577	0.0000	1.0000
Translate Diagram	0.1647	0.3731	0.0000	1.0000
Visualization	0.0824	0.2765	0.0000	1.0000
Semantic Memory	1.5765	0.9306	1.0000	4.0000
Solution Planning				
Presence of Subgoals	0.0941	0.2937	0.0000	1.0000
Number of Subgoals	0.1412	0.5379	0.0000	4.0000
Relative Definition of Variables	0.0235	0.1525	0.0000	1.0000
Solution Execution				
Number Knowledge	3.0824	1.7404	0.0000	5.0000
1. Single-digit	0.6588	0.4769	0.0000	1.0000
2. Double-digit	0.5412	0.5013	0.0000	1.0000
3. Triple-digit	0.1176	0.3241	0.0000	1.0000
4. Four-digit +	0.0941	0.2937	0.0000	1.0000
5. Fraction/Decimal	0.3882	0.4902	0.0000	1.0000
Alt. Procedural Knowledge	1.0118	0.9322	0.0000	3.0000
1. Multiple Steps	0.4235	0.4971	0.0000	1.0000
2. Algebraic Equations	0.3412	0.4769	0.0000	1.0000
3. Mixed fractions	0.0353	0.1856	0.0000	1.0000
Procedural Knowledge	1.9647	2.1013	0.0000	6.0000
1. Integers	0.2471	0.4339	0.0000	1.0000
2. Fractions	0.2706	0.4469	0.0000	1.0000
3. Proportions	0.0824	0.2765	0.0000	1.0000
4. Decimals	0.1059	0.3095	0.0000	1.0000
5. Negative Numbers	0.1294	0.3376	0.0000	1.0000
6. Square Roots	0.0824	0.2765	0.0000	1.0000
Number of Procedures	0.8824	0.8648	0.0000	3.0000
Number of Computations	2.0118	2.6793	0.0000	16.0000
Number of Operands	1.4000	1.0142	0.0000	5.0000
Meta-Cognition Process	0.4588	0.5013	0.0000	1.0000
Decision Processing				
Decision Processing Confirmation	0.5059	0.5029	0.0000	1.0000

	Mean	SD	Min	Max
Bottom-Up Processing	0.2118	0.4110	0.0000	1.0000
Top-Down Processing	0.2941	0.4583	0.0000	1.0000
Functional Distractors	2.0471	1.0340	0.0000	3.0000

Table B 3. *Cognitive variable scores for translation attributes.*

Item	Total	Mathematical	Contextual	Stem	Content Words
p1i1	41	21	20	21	19
p1i2	39	10	29	27	22
p1i3	58	5	53	50	40
p1i5	74	32	42	49	29
p1i6	51	14	37	31	22
p1i7	20	0	20	11	7
p1i8	26	17	9	22	18
p1i9	42	4	38	16	13
p1i10	37	1	36	17	9
p1i11	32	24	8	12	9
p1i12	42	3	39	18	11
p1i13	47	8	39	43	29
p1i14	29	15	14	25	18
p1i15	38	19	19	26	17
p1i16	72	15	57	60	50
p1i17	86	12	74	46	36
p1i18	91	16	75	40	28
p1i19	50	17	33	34	25
p1i20	43	16	27	35	25
p1i21	42	14	28	34	25
p1i22	47	20	27	23	18
p1i23	62	17	45	54	38
p1i24	83	34	49	35	27
p1i25	80	34	46	60	47
p1i26	171	97	74	65	46
p1i27	117	44	73	89	73
p1i28	86	38	48	60	47
p1i29	130	61	69	82	68
p1i30	83	44	39	55	42
p2i1	64	15	49	56	40
p2i2	28	20	8	16	12
p2i3	72	29	43	47	28
p2i4	36	28	8	16	12
p2i5	69	14	55	61	43
p2i6	117	1	116	39	29
p2i7	61	13	48	53	37
p2i8	63	19	44	35	26

Item	Total	Mathematical	Contextual	Stem	Content Words
p2i9	42	8	34	34	23
p2i10	49	10	39	41	28
p2i11	42	9	33	34	20
p2i12	44	8	36	40	28
p2i13	34	12	22	26	21
p2i14	56	29	27	40	30
p2i15	64	16	48	56	37
p2i16	61	10	51	53	44
p2i17	69	17	52	61	36
p2i18	34	7	27	30	17
p2i19	33	6	27	29	17
p2i20	54	13	41	46	28
p2i21	66	32	34	38	32
p2i22	36	12	24	28	18
p2i23	32	10	22	24	14
p2i24	64	18	46	56	45
p2i25	72	18	54	60	42
p2i26	27	10	17	17	14
p2i27	45	20	25	33	25
p3i1	56	19	37	40	28
p3i2	59	27	32	35	28
p3i3	82	32	50	54	36
p3i4	69	32	37	41	27
p3i5	62	31	31	34	24
p3i6	67	24	43	47	35
p3i7	87	32	55	59	47
p3i8	36	26	10	24	21
p3i9	12	7	5	8	5
p3i10	34	4	30	22	15
p3i11	25	18	7	21	17
p3i12	12	7	5	8	5
p3i13	23	16	7	19	15
p3i14	20	12	8	8	5
p3i15	20	0	20	12	6
p3i16	17	1	16	9	7
p3i17	24	4	20	18	13
p3i18	54	39	15	46	40
p3i19	39	0	39	13	7
p3i20	36	19	17	24	16

Item	Total	Mathematical	Contextual	Stem	Content Words
p3i21	76	16	60	37	27
p3i22	50	38	12	42	36
p3i23	44	30	14	40	33
p3i24	46	36	10	34	30
p3i25	50	30	20	38	28
p3i26	32	14	18	20	16
p3i27	97	43	54	61	45
p3i28	107	27	80	57	47
p3i29	90	55	35	41	33

Table B 4. *Cognitive variable scores for text comprehension variables.*

Item	Flesch Reading Ease Test	Flesch-Kincaid Grade Level Test	Average - LSA	Comp. One	Comp. Two	Comp. Three	Comp. Four
p1i1	66.7	5.6	0.44	0.86	0.02		
p1i2	99.1	2.7	0.435	0.45	0.42		
p1i3	85	5.5	0.66	0.7	0.62		
p1i5	59.5	8.8	0.75	0.66	0.84		
p1i6	98.1	2.1	0.35	0.31	0.39		
p1i7	57.2	8	1	1			
p1i8	95.7	3.6	1	1			
p1i9	55.9	7.4	-0.05	-0.05			
p1i10	60.1	9	1	1			
p1i11	75.5	4.9	1	1			
p1i12	52	8.2	0.18	0.18			
p1i13	80.3	6.4	0.19	0.19			
p1i14	97	1.4	0.81	0.81			
p1i15	52	8.2	0.16	0.16			
p1i16	56.2	9.1	0.67	0.67			
p1i17	79.9	4.1	0.04	0.05	0.03		
p1i18	72.3	8.1	0.09	0.09			
p1i19	73.7	5.8	0.44	0.44			
p1i20	82.4	4.2	0.39	0.39			
p1i21	82.4	4.8	0.63	0.63			
p1i22	71.7	6.7	1	1			
p1i23	72.3	8.1	0.69	0.69			
p1i24	56.1	9.7	0.28	0.28			
p1i25	76.6	6.1	0.41667	0.48	0.34	0.43	
p1i26	56.2	9.1	0.41	0.68	0.34	0.21	
p1i27	61.7	5.2	0.4325	0.51	0.55	0.37	0.3
p1i28	65.1	8.8	0.49	0.49			
p1i29	72.7	6.9	0.505	0.6	0.5	0.23	0.69
p1i30	51.9	11.1	0.6	0.6			
p2i1	83.9	4.5	0.36	0.17	0.17	0.74	
p2i2	87.9	3.7	1	1			
p2i3	63.1	8.3	0.465	0.8	0.13		
p2i4	95.9	2.8	1	1			
p2i5	80.4	6.5	0.46	0.65	0.27		

Item	Flesch Reading Ease Test	Flesch-Kincaid Grade Level Test	Average - LSA	Comp. One	Comp. Two	Comp. Three	Comp. Four
p2i6	93.8	3.3	0.3	0.22	0.38		
p2i7	68.2	6.5	0.22333	0.34	0.22	0.11	
p2i8	76.5	5.8	0.22	0.22			
p2i9	43.5	9.7	0.065	0.12	0.01		
p2i10	62.8	9.2	0.34	0.34			
p2i11	79.2	6	0.25	0.25			
p2i12	66.2	7	0.625	0.81	0.44		
p2i13	63.4	7.6	0.81	0.81			
p2i14	90.1	4.7	0.38	0.38			
p2i15	69.9	6.7	0.62333	0.7	0.47	0.7	
p2i16	52.7	8.8	0.5	0.43	0.49	0.58	
p2i17	85.6	4.9	0.42	0.28	0.33	0.65	
p2i18	83.2	4	0.845	0.77	0.92		
p2i19	92	2.8	0.4	-0.04	0.84		
p2i20	66.6	9.1	0.2	0.24	0.16		
p2i21	76.2	7.3	0.25	0.25			
p2i22	74.7	6.3	0.72	0.72			
p2i23	50.1	9.2	0.57	0.57			
p2i24	69.3	6.4	0.535	0.59	0.48		
p2i25	74	6.2	0.64667	0.73	0.81	0.4	
p2i26	88.7	2.9	0.33	0.33			
p2i27	66.4	6.8	0.19	0.19			
p3i1	76.5	7.5	0.42	0.42			
p3i2	62	7.5	0.13	-0.01	0.33	0.07	
p3i3	66.3	8.4	0.69	0.72	0.66		
p3i4	68.5	7	0.33	0.51	0.15		
p3i5	70	6.2	0.07	0.06	0.08		
p3i6	87.8	4.4	0.4	0.48	0.32		
p3i7	83	4.6	0.2	0.23	0.15	0.23	0.19
p3i8	68.9	7.6	1	1			
p3i9	42.6	9	1	1			
p3i10	83.7	3.6	0.22	0.42	0.02		
p3i11	78.8	3.9	1	1			
p3i12	100	0.6	1	1			
p3i13	89.5	4	1	1			
p3i14	61.2	6.7	1	1			

Item	Flesch Reading Ease Test	Flesch-Kincaid Grade Level Test	Average - LSA	Comp. One	Comp. Two	Comp. Three	Comp. Four
p3i15	67.7	6.7	1	1			
p3i16	66.1	6.2	1	1			
p3i17	59.6	9.9	1	1			
p3i18	93.5	1.5	0.2	0.2			
p3i19	96	3	1	1			
p3i20	73.7	5	0.29	0.29			
p3i21	87.3	4.9	0.8	0.8			
p3i22	100	0	0.92	0.92			
p3i23	100	0	0.31	0.31			
p3i24	100	0	0.93	0.93			
p3i25	78.2	4.8	0.67	0.67			
p3i26	65.5	6.6	0.76	0.76			
p3i27	66.9	8.9	0.69	0.73	0.65		
p3i28	68.3	6.7	0.17	0.26	0.14	0.11	
p3i29	73.2	6.4	0.36	0.49	0.23		

Table B 5. *Cognitive variable scores for mathematical propositions, translation word equation, and encode diagram.*

Item	Assignment Propositions	Relation Propositions	Translate Word Equation	Encode Diagram
p1i1	1	0	1	0
p1i2	1	0	1	0
p1i3	0	0	1	0
p1i5	0	0	0	0
p1i6	1	0	1	0
p1i7	0	0	0	0
p1i8	1	0	0	0
p1i9	0	0	0	0
p1i10	0	0	0	0
p1i11	0	0	1	0
p1i12	0	0	0	0
p1i13	0	0	0	0
p1i14	0	0	0	0
p1i15	0	0	0	0
p1i16	0	0	0	0
p1i17	0	0	0	0
p1i18	0	0	0	0
p1i19	0	0	0	1
p1i20	0	0	0	1
p1i21	0	0	1	0
p1i22	0	0	0	0
p1i23	0	0	0	1
p1i24	0	0	1	0
p1i25	0	0	1	0
p1i26	0	0	0	1
p1i27	0	0	1	0
p1i28	0	0	1	0
p1i29	0	0	0	0
p1i30	0	0	1	0
p2i1	1	0	0	0
p2i2	0	0	0	0
p2i3	0	1	1	0
p2i4	0	0	0	0
p2i5	0	0	0	0

Item	Assignment Propositions	Relation Propositions	Translate Word Equation	Encode Diagram
p2i6	0	0	0	0
p2i7	0	0	0	1
p2i8	0	0	0	1
p2i9	0	0	0	0
p2i10	0	0	0	0
p2i11	0	0	0	1
p2i12	0	0	0	0
p2i13	0	1	0	0
p2i14	0	0	0	1
p2i15	1	0	0	1
p2i16	1	0	0	1
p2i17	1	1	0	1
p2i18	0	0	0	0
p2i19	0	0	0	0
p2i20	0	0	0	1
p2i21	1	0	1	0
p2i22	0	0	0	0
p2i23	0	0	0	0
p2i24	0	0	1	0
p2i25	0	0	1	0
p2i26	0	0	1	0
p2i27	0	0	1	1
p3i1	1	0	1	0
p3i2	1	0	1	0
p3i3	1	0	1	0
p3i4	1	0	1	0
p3i5	1	0	1	0
p3i6	1	0	0	1
p3i7	1	0	1	0
p3i8	0	0	0	0
p3i9	0	0	0	0
p3i10	0	0	0	0
p3i11	0	0	0	0
p3i12	0	0	0	0
p3i13	0	0	0	0
p3i14	0	0	0	0
p3i15	0	0	0	0

Item	Assignment Propositions	Relation Propositions	Translate Word Equation	Encode Diagram
p3i16	0	0	0	0
p3i17	0	0	0	0
p3i18	0	0	0	1
p3i19	0	0	0	0
p3i20	0	0	0	1
p3i21	0	1	0	0
p3i22	0	0	0	1
p3i23	0	0	0	1
p3i24	0	0	0	1
p3i25	0	0	0	1
p3i26	1	0	1	0
p3i27	1	0	1	0
p3i28	0	0	1	0
p3i29	0	0	1	0

Table B 6. *Cognitive variable scores for contextual propositions, part 1.*

Item	Total Number of Propositions	Total Proposition Density	Number of Predicate Propositions	Predicate Density	Number of Modifier Propositions	Modifier Density
p1i1	9	0.4286	4	0.1905	3	0.1429
p1i2	10	0.3704	3	0.1111	5	0.1852
p1i3	20	0.4000	10	0.2000	7	0.1400
p1i5	11	0.2245	5	0.1020	5	0.1020
p1i6	10	0.3226	5	0.1613	3	0.0968
p1i7	2	0.1818	1	0.0909	1	0.0909
p1i8	6	0.2727	2	0.0909	2	0.0909
p1i9	6	0.3750	3	0.1875	2	0.1250
p1i10	4	0.2353	3	0.1765	0	0.0000
p1i11	2	0.1667	2	0.1667	0	0.0000
p1i12	7	0.3889	3	0.1667	4	0.2222
p1i13	9	0.2093	2	0.0465	6	0.1395
p1i14	3	0.1200	1	0.0400	2	0.0800
p1i15	4	0.1538	2	0.0769	2	0.0769
p1i16	12	0.2000	2	0.0333	10	0.1667
p1i17	8	0.1739	4	0.0870	4	0.0870
p1i18	10	0.2500	3	0.0750	6	0.1500
p1i19	7	0.2059	2	0.0588	5	0.1471
p1i20	6	0.1714	1	0.0286	5	0.1429
p1i21	11	0.3235	1	0.0294	9	0.2647
p1i22	6	0.2609	1	0.0435	5	0.2174
p1i23	14	0.2593	3	0.0556	10	0.1852
p1i24	14	0.4000	7	0.2000	6	0.1714
p1i25	20	0.3333	9	0.1500	10	0.1667
p1i26	19	0.2923	8	0.1231	9	0.1385
p1i27	30	0.3371	13	0.1461	14	0.1573
p1i28	12	0.2000	2	0.0333	7	0.1167
p1i29	30	0.3659	14	0.1707	15	0.1829
p1i30	13	0.2364	4	0.0727	7	0.1273
p2i1	15	0.2679	8	0.1429	6	0.1071
p2i2	3	0.1875	2	0.1250	1	0.0625
p2i3	17	0.3617	5	0.1064	12	0.2553
p2i4	3	0.1875	2	0.1250	1	0.0625
p2i5	20	0.3279	4	0.0656	14	0.2295
p2i6	12	0.3077	4	0.1026	8	0.2051
p2i7	13	0.2453	5	0.0943	8	0.1509

Item	Total Number of Propositions	Total Proposition Density	Number of Predicate Propositions	Predicate Density	Number of Modifier Propositions	Modifier Density
p2i8	8	0.2286	4	0.1143	4	0.1143
p2i9	8	0.2353	4	0.1176	4	0.1176
p2i10	11	0.2683	6	0.1463	5	0.1220
p2i11	11	0.3235	3	0.0882	8	0.2353
p2i12	12	0.3000	7	0.1750	4	0.1000
p2i13	10	0.3846	3	0.1154	7	0.2692
p2i14	14	0.3500	3	0.0750	9	0.2250
p2i15	17	0.3036	6	0.1071	11	0.1964
p2i16	17	0.3208	5	0.0943	8	0.1509
p2i17	19	0.3115	5	0.0820	13	0.2131
p2i18	8	0.2667	5	0.1667	3	0.1000
p2i19	7	0.2414	3	0.1034	4	0.1379
p2i20	16	0.3478	9	0.1957	7	0.1522
p2i21	15	0.3947	3	0.0789	10	0.2632
p2i22	10	0.3571	2	0.0714	7	0.2500
p2i23	10	0.4167	2	0.0833	8	0.3333
p2i24	15	0.2679	6	0.1071	7	0.1250
p2i25	17	0.2833	9	0.1500	7	0.1167
p2i26	7	0.4118	3	0.1765	4	0.2353
p2i27	8	0.2424	3	0.0909	5	0.1515
p3i1	12	0.3000	6	0.1500	5	0.1250
p3i2	16	0.4571	4	0.1143	9	0.2571
p3i3	17	0.3148	8	0.1481	8	0.1481
p3i4	11	0.2683	7	0.1707	3	0.0732
p3i5	13	0.3824	5	0.1471	7	0.2059
p3i6	14	0.2979	6	0.1277	7	0.1489
p3i7	24	0.4068	12	0.2034	12	0.2034
p3i8	3	0.1250	2	0.0833	1	0.0417
p3i9	1	0.1250	1	0.1250	0	0.0000
p3i10	8	0.3636	1	0.0455	5	0.2273
p3i11	2	0.0952	2	0.0952	0	0.0000
p3i12	1	0.1250	1	0.1250	0	0.0000
p3i13	2	0.1053	2	0.1053	0	0.0000
p3i14	3	0.3750	1	0.1250	2	0.2500
p3i15	6	0.5000	2	0.1667	3	0.2500
p3i16	3	0.3333	1	0.1111	2	0.2222
p3i17	8	0.4444	2	0.1111	5	0.2778

Item	Total Number of Propositions	Total Proposition Density	Number of Predicate Propositions	Predicate Density	Number of Modifier Propositions	Modifier Density
p3i18	3	0.0652	2	0.0435	1	0.0217
p3i19	3	0.2308	1	0.0769	2	0.1538
p3i20	5	0.2083	1	0.0417	4	0.1667
p3i21	10	0.2703	5	0.1351	4	0.1081
p3i22	3	0.0714	1	0.0238	2	0.0476
p3i23	4	0.1000	2	0.0500	2	0.0500
p3i24	4	0.1176	2	0.0588	2	0.0588
p3i25	6	0.1579	2	0.0526	4	0.1053
p3i26	9	0.4500	4	0.2000	3	0.1500
p3i27	25	0.4098	7	0.1148	17	0.2787
p3i28	20	0.3509	11	0.1930	8	0.1404
p3i29	15	0.3659	7	0.1707	5	0.1220

Table B 7. *Cognitive variable scores for contextual propositions, part 2.*

Item	Number of Connector Propositions	Connector Density	Total Number of Unique Arguments	Unique Argument Density
p1i1	2	0.0952	9	0.4286
p1i2	2	0.0741	9	0.3333
p1i3	3	0.0600	16	0.3200
p1i5	1	0.0204	9	0.1837
p1i6	2	0.0645	8	0.2581
p1i7	0	0.0000	4	0.3636
p1i8	2	0.0909	7	0.3182
p1i9	1	0.0625	7	0.4375
p1i10	1	0.0588	5	0.2941
p1i11	0	0.0000	3	0.2500
p1i12	0	0.0000	8	0.4444
p1i13	1	0.0233	12	0.2791
p1i14	0	0.0000	4	0.1600
p1i15	0	0.0000	6	0.2308
p1i16	0	0.0000	12	0.2000
p1i17	0	0.0000	9	0.1957
p1i18	1	0.0250	12	0.3000
p1i19	0	0.0000	8	0.2353
p1i20	0	0.0000	7	0.2000
p1i21	1	0.0294	10	0.2941
p1i22	0	0.0000	7	0.3043
p1i23	1	0.0185	12	0.2222
p1i24	1	0.0286	14	0.4000
p1i25	1	0.0167	17	0.2833
p1i26	2	0.0308	17	0.2615
p1i27	3	0.0337	25	0.2809
p1i28	3	0.0500	12	0.2000
p1i29	1	0.0122	24	0.2927
p1i30	2	0.0364	13	0.2364
p2i1	1	0.0179	12	0.2143
p2i2	0	0.0000	4	0.2500
p2i3	0	0.0000	15	0.3191
p2i4	0	0.0000	4	0.2500
p2i5	2	0.0328	14	0.2295
p2i6	0	0.0000	12	0.3077

Item	Number of Connector Propositions	Connector Density	Total Number of Unique Arguments	Unique Argument Density
p2i7	0	0.0000	14	0.2642
p2i8	0	0.0000	9	0.2571
p2i9	0	0.0000	9	0.2647
p2i10	0	0.0000	11	0.2683
p2i11	0	0.0000	10	0.2941
p2i12	1	0.0250	9	0.2250
p2i13	0	0.0000	9	0.3462
p2i14	2	0.0500	10	0.2500
p2i15	0	0.0000	15	0.2679
p2i16	4	0.0755	14	0.2642
p2i17	1	0.0164	16	0.2623
p2i18	0	0.0000	7	0.2333
p2i19	0	0.0000	8	0.2759
p2i20	0	0.0000	16	0.3478
p2i21	2	0.0526	14	0.3684
p2i22	1	0.0357	10	0.3571
p2i23	0	0.0000	9	0.3750
p2i24	2	0.0357	15	0.2679
p2i25	1	0.0167	12	0.2000
p2i26	0	0.0000	7	0.4118
p2i27	0	0.0000	9	0.2727
p3i1	1	0.0250	12	0.3000
p3i2	3	0.0857	15	0.4286
p3i3	1	0.0185	15	0.2778
p3i4	1	0.0244	10	0.2439
p3i5	1	0.0294	14	0.4118
p3i6	1	0.0213	15	0.3191
p3i7	0	0.0000	22	0.3729
p3i8	0	0.0000	4	0.1667
p3i9	0	0.0000	2	0.2500
p3i10	2	0.0909	10	0.4545
p3i11	0	0.0000	3	0.1429
p3i12	0	0.0000	2	0.2500
p3i13	0	0.0000	3	0.1579
p3i14	0	0.0000	4	0.5000
p3i15	1	0.0833	5	0.4167
p3i16	0	0.0000	4	0.4444

Item	Number of Connector Propositions	Connector Density	Total Number of Unique Arguments	Unique Argument Density
p3i17	1	0.0556	5	0.2778
p3i18	0	0.0000	5	0.1087
p3i19	0	0.0000	3	0.2308
p3i20	0	0.0000	5	0.2083
p3i21	1	0.0270	11	0.2973
p3i22	0	0.0000	4	0.0952
p3i23	0	0.0000	5	0.1250
p3i24	0	0.0000	5	0.1471
p3i25	0	0.0000	7	0.1842
p3i26	2	0.1000	8	0.4000
p3i27	1	0.0164	18	0.2951
p3i28	1	0.0175	18	0.3158
p3i29	3	0.0732	18	0.4390

Table B 8. *Cognitive variable scores for contextual propositions, part 3.*

Item	Total Number of Arguments	Total Argument Ratio	Max. Number of Arguments	Relevant Propositions	Density of Relevant Propositions
p1i1	24	1.1429	5	9	1.0000
p1i2	33	1.2222	5	7	0.7000
p1i3	81	1.6200	13	20	1.0000
p1i5	32	0.6531	5	10	0.9091
p1i6	30	0.9677	6	9	0.9000
p1i7	7	0.6364	4	2	1.0000
p1i8	20	0.9091	7	6	1.0000
p1i9	19	1.1875	6	6	1.0000
p1i10	12	0.7059	5	4	1.0000
p1i11	5	0.4167	3	2	1.0000
p1i12	24	1.3333	6	7	1.0000
p1i13	31	0.7209	6	6	0.6667
p1i14	8	0.3200	3	3	1.0000
p1i15	12	0.4615	4	4	1.0000
p1i16	49	0.8167	8	12	1.0000
p1i17	23	0.5000	5	8	1.0000
p1i18	32	0.8000	9	8	0.8000
p1i19	25	0.7353	6	7	1.0000
p1i20	18	0.5143	4	6	1.0000
p1i21	45	1.3235	12	11	1.0000
p1i22	20	0.8696	7	6	1.0000
p1i23	57	1.0556	9	14	1.0000
p1i24	62	1.7714	10	14	1.0000
p1i25	90	1.5000	12	10	0.5000
p1i26	63	0.9692	9	10	0.5263
p1i27	129	1.4494	13	11	0.3667
p1i28	59	0.9833	10	11	0.9167
p1i29	133	1.6220	16	12	0.4000
p1i30	50	0.9091	11	13	1.0000
p2i1	51	0.9107	8	5	0.3333
p2i2	8	0.5000	4	3	1.0000
p2i3	63	1.3404	9	17	1.0000
p2i4	8	0.5000	4	3	1.0000
p2i5	79	1.2951	15	20	1.0000
p2i6	39	1.0000	6	12	1.0000
p2i7	41	0.7736	5	9	0.6923

Item	Total Number of Arguments	Total Argument Ratio	Max. Number of Arguments	Relevant Propositions	Density of Relevant Propositions
p2i8	27	0.7714	8	8	1.0000
p2i9	21	0.6176	5	8	1.0000
p2i10	36	0.8780	7	11	1.0000
p2i11	37	1.0882	8	11	1.0000
p2i12	39	0.9750	7	12	1.0000
p2i13	32	1.2308	7	10	1.0000
p2i14	71	1.7750	12	14	1.0000
p2i15	62	1.1071	7	17	1.0000
p2i16	64	1.2075	7	1	0.0588
p2i17	71	1.1639	9	15	0.7895
p2i18	24	0.8000	5	8	1.0000
p2i19	21	0.7241	4	7	1.0000
p2i20	55	1.1957	7	14	0.8750
p2i21	68	1.7895	11	15	1.0000
p2i22	36	1.2857	9	9	0.9000
p2i23	31	1.2917	6	10	1.0000
p2i24	57	1.0179	7	15	1.0000
p2i25	64	1.0667	7	17	1.0000
p2i26	22	1.2941	5	7	1.0000
p2i27	29	0.8788	6	8	1.0000
p3i1	51	1.2750	9	12	1.0000
p3i2	54	1.5429	9	16	1.0000
p3i3	67	1.2407	8	17	1.0000
p3i4	35	0.8537	7	11	1.0000
p3i5	45	1.3235	8	13	1.0000
p3i6	53	1.1277	8	14	1.0000
p3i7	93	1.5763	10	24	1.0000
p3i8	9	0.3750	4	3	1.0000
p3i9	2	0.2500	2	1	1.0000
p3i10	24	1.0909	6	8	1.0000
p3i11	5	0.2381	3	2	1.0000
p3i12	2	0.2500	2	1	1.0000
p3i13	5	0.2632	3	2	1.0000
p3i14	9	1.1250	4	3	1.0000
p3i15	31	2.5833	9	6	1.0000
p3i16	8	0.8889	4	3	1.0000
p3i17	30	1.6667	8	8	1.0000

Item	Total Number of Arguments	Total Argument Ratio	Max. Number of Arguments	Relevant Propositions	Density of Relevant Propositions
p3i18	7	0.1522	3	3	1.0000
p3i19	8	0.6154	3	3	1.0000
p3i20	13	0.5417	3	5	1.0000
p3i21	35	0.9459	6	10	1.0000
p3i22	7	0.1667	3	3	1.0000
p3i23	11	0.2750	4	4	1.0000
p3i24	12	0.3529	4	4	1.0000
p3i25	19	0.5000	5	6	1.0000
p3i26	28	1.4000	6	7	0.7778
p3i27	146	2.3934	18	25	1.0000
p3i28	96	1.6842	13	20	1.0000
p3i29	66	1.6098	10	14	0.9333

Table B 9. *Cognitive variable scores for contextual propositions, part 4.*

Item	Relevant Words	Density of Relevant Words	Irrelevant Propositions	Density of Irrelevant Propositions	Irrelevant Words	Density of Irrelevant Words
p1i1	21	1.0000	0	0.0000	0	0.0000
p1i2	20	0.7407	3	0.3000	7	0.2593
p1i3	50	1.0000	0	0.0000	0	0.0000
p1i5	48	0.9796	1	0.0909	1	0.0204
p1i6	28	0.9032	1	0.1000	3	0.0968
p1i7	11	1.0000	0	0.0000	0	0.0000
p1i8	22	1.0000	0	0.0000	0	0.0000
p1i9	16	1.0000	0	0.0000	0	0.0000
p1i10	17	1.0000	0	0.0000	0	0.0000
p1i11	12	1.0000	0	0.0000	0	0.0000
p1i12	18	1.0000	0	0.0000	0	0.0000
p1i13	33	0.7674	3	0.3333	10	0.2326
p1i14	25	1.0000	0	0.0000	0	0.0000
p1i15	26	1.0000	0	0.0000	0	0.0000
p1i16	60	1.0000	0	0.0000	0	0.0000
p1i17	46	1.0000	0	0.0000	0	0.0000
p1i18	38	0.9500	2	0.2000	2	0.0500
p1i19	34	1.0000	0	0.0000	0	0.0000
p1i20	35	1.0000	0	0.0000	0	0.0000
p1i21	34	1.0000	0	0.0000	0	0.0000
p1i22	23	1.0000	0	0.0000	0	0.0000
p1i23	54	1.0000	0	0.0000	0	0.0000
p1i24	35	1.0000	0	0.0000	0	0.0000
p1i25	36	0.6000	10	0.5000	24	0.4000
p1i26	37	0.5692	9	0.4737	28	0.4308
p1i27	47	0.5281	19	0.6333	42	0.4719
p1i28	60	1.0000	1	0.0833	0	0.0000
p1i29	50	0.6098	18	0.6000	32	0.3902
p1i30	55	1.0000	0	0.0000	0	0.0000
p2i1	33	0.5893	10	0.6667	23	0.4107
p2i2	16	1.0000	0	0.0000	0	0.0000
p2i3	47	1.0000	0	0.0000	0	0.0000
p2i4	16	1.0000	0	0.0000	0	0.0000
p2i5	61	1.0000	0	0.0000	0	0.0000
p2i6	39	1.0000	0	0.0000	0	0.0000

Item	Relevant Words	Density of Relevant Words	Irrelevant Propositions	Density of Irrelevant Propositions	Irrelevant Words	Density of Irrelevant Words
p2i7	43	0.8113	4	0.3077	10	0.1887
p2i8	35	1.0000	0	0.0000	0	0.0000
p2i9	34	1.0000	0	0.0000	0	0.0000
p2i10	41	1.0000	0	0.0000	0	0.0000
p2i11	34	1.0000	0	0.0000	0	0.0000
p2i12	40	1.0000	0	0.0000	0	0.0000
p2i13	26	1.0000	0	0.0000	0	0.0000
p2i14	40	1.0000	0	0.0000	0	0.0000
p2i15	56	1.0000	0	0.0000	0	0.0000
p2i16	15	0.2830	16	0.9412	38	0.7170
p2i17	50	0.8197	4	0.2105	11	0.1803
p2i18	30	1.0000	0	0.0000	0	0.0000
p2i19	29	1.0000	0	0.0000	0	0.0000
p2i20	39	0.8478	2	0.1250	7	0.1522
p2i21	38	1.0000	0	0.0000	0	0.0000
p2i22	27	0.9643	1	0.1000	1	0.0357
p2i23	24	1.0000	0	0.0000	0	0.0000
p2i24	56	1.0000	0	0.0000	0	0.0000
p2i25	60	1.0000	0	0.0000	0	0.0000
p2i26	17	1.0000	0	0.0000	0	0.0000
p2i27	33	1.0000	0	0.0000	0	0.0000
p3i1	40	1.0000	0	0.0000	0	0.0000
p3i2	35	1.0000	0	0.0000	0	0.0000
p3i3	54	1.0000	0	0.0000	0	0.0000
p3i4	41	1.0000	0	0.0000	0	0.0000
p3i5	34	1.0000	0	0.0000	0	0.0000
p3i6	47	1.0000	0	0.0000	0	0.0000
p3i7	59	1.0000	0	0.0000	0	0.0000
p3i8	24	1.0000	0	0.0000	0	0.0000
p3i9	8	1.0000	0	0.0000	0	0.0000
p3i10	22	1.0000	0	0.0000	0	0.0000
p3i11	21	1.0000	0	0.0000	0	0.0000
p3i12	8	1.0000	0	0.0000	0	0.0000
p3i13	19	1.0000	0	0.0000	0	0.0000
p3i14	8	1.0000	0	0.0000	0	0.0000
p3i15	12	1.0000	0	0.0000	0	0.0000
p3i16	9	1.0000	0	0.0000	0	0.0000

Item	Relevant Words	Density of Relevant Words	Irrelevant Propositions	Density of Irrelevant Propositions	Irrelevant Words	Density of Irrelevant Words
p3i17	18	1.0000	0	0.0000	0	0.0000
p3i18	46	1.0000	0	0.0000	0	0.0000
p3i19	13	1.0000	0	0.0000	0	0.0000
p3i20	24	1.0000	0	0.0000	0	0.0000
p3i21	37	1.0000	0	0.0000	0	0.0000
p3i22	42	1.0000	0	0.0000	0	0.0000
p3i23	40	1.0000	0	0.0000	0	0.0000
p3i24	34	1.0000	0	0.0000	0	0.0000
p3i25	38	1.0000	0	0.0000	0	0.0000
p3i26	14	0.7000	2	0.2222	6	0.3000
p3i27	61	1.0000	0	0.0000	0	0.0000
p3i28	57	1.0000	0	0.0000	0	0.0000
p3i29	39	0.9512	1	0.0667	2	0.0488

Table B 10. *Cognitive variable scores for integration component.*

Item	Given Equation - in Stem	Generate Eq. or Possible Values	Access Equation	Auxiliary Diagram	Translate Diagram	Visualization
p1i1	0	0	0	0	0	0
p1i2	0	0	0	0	0	0
p1i3	0	0	0	0	0	0
p1i5	1	0	0	0	0	0
p1i6	0	1	0	0	0	0
p1i7	0	0	0	0	0	0
p1i8	1	0	0	0	0	0
p1i9	0	1	0	0	0	0
p1i10	0	0	0	0	0	0
p1i11	0	0	0	0	0	0
p1i12	0	1	0	0	0	0
p1i13	0	0	1	0	0	0
p1i14	0	0	1	0	0	0
p1i15	0	0	1	0	0	0
p1i16	0	0	1	0	0	0
p1i17	0	0	0	0	0	0
p1i18	0	0	1	0	0	0
p1i19	1	1	0	0	1	0
p1i20	1	0	0	0	1	0
p1i21	1	0	0	0	0	1
p1i22	1	0	0	0	0	1
p1i23	1	0	0	1	0	0
p1i24	0	1	0	0	0	0
p1i25	0	1	0	0	0	0
p1i26	1	0	0	0	1	0
p1i27	0	1	0	0	0	0
p1i28	0	1	0	0	0	0
p1i29	1	0	0	0	0	0
p1i30	0	1	0	0	0	0
p2i1	1	0	0	0	0	0
p2i2	1	0	0	0	0	0
p2i3	0	1	0	0	0	0
p2i4	1	0	0	0	0	0
p2i5	1	0	0	0	0	0
p2i6	0	0	0	0	0	0

Item	Given Equation - in Stem	Generate Eq. or Possible Values	Access Equation	Auxiliary Diagram	Translate Diagram	Visualization
p2i7	0	1	0	1	0	0
p2i8	0	1	0	0	1	0
p2i9	0	1	0	0	0	0
p2i10	0	1	0	0	0	0
p2i11	0	1	0	0	1	0
p2i12	0	1	0	0	0	0
p2i13	0	1	0	0	0	1
p2i14	0	1	0	1	0	0
p2i15	0	1	0	0	1	0
p2i16	0	1	0	1	0	0
p2i17	0	1	0	1	0	0
p2i18	0	1	0	0	0	0
p2i19	0	1	0	0	0	0
p2i20	0	0	1	1	0	0
p2i21	0	0	0	0	0	0
p2i22	0	0	1	0	0	1
p2i23	0	0	1	0	0	1
p2i24	0	0	0	0	0	0
p2i25	0	0	0	0	0	0
p2i26	0	0	0	0	0	0
p2i27	0	0	0	0	1	0
p3i1	0	1	0	0	0	0
p3i2	0	1	0	0	0	0
p3i3	0	1	0	0	0	0
p3i4	0	1	0	0	0	0
p3i5	0	1	0	0	0	0
p3i6	0	1	0	0	1	0
p3i7	0	1	0	0	0	0
p3i8	1	0	0	0	0	0
p3i9	1	0	0	0	0	0
p3i10	0	1	0	0	0	0
p3i11	1	0	0	0	0	0
p3i12	1	0	0	0	0	0
p3i13	1	0	0	0	0	0
p3i14	0	0	0	0	0	0
p3i15	0	0	0	0	0	0
p3i16	0	0	0	0	0	0

Item	Given Equation - in Stem	Generate Eq. or Possible Values	Access Equation	Auxiliary Diagram	Translate Diagram	Visualization
p3i17	0	0	0	0	0	0
p3i18	0	0	1	0	1	0
p3i19	0	0	0	0	0	0
p3i20	0	0	0	0	1	0
p3i21	1	0	0	0	0	1
p3i22	0	0	1	0	1	0
p3i23	0	0	1	0	1	0
p3i24	0	0	1	0	1	0
p3i25	0	0	0	0	1	0
p3i26	0	1	0	0	0	0
p3i27	0	1	0	0	0	1
p3i28	1	1	0	0	0	0
p3i29	0	1	0	0	0	0

Table B 11. *Cognitive variable scores for solution planning component.*

Item	Presence of Subgoals	Number of Subgoals	Relative Definition of Variables
p1i1	0	0	0
p1i2	0	0	0
p1i3	0	0	0
p1i5	0	0	0
p1i6	0	0	0
p1i7	0	0	0
p1i8	0	0	0
p1i9	0	0	0
p1i10	0	0	0
p1i11	0	0	0
p1i12	0	0	0
p1i13	1	4	0
p1i14	0	0	0
p1i15	0	0	0
p1i16	0	0	0
p1i17	0	0	0
p1i18	0	0	0
p1i19	0	0	0
p1i20	0	0	0
p1i21	0	0	0
p1i22	0	0	0
p1i23	0	0	0
p1i24	1	1	0
p1i25	0	0	0
p1i26	0	0	0
p1i27	1	1	0
p1i28	1	2	0
p1i29	0	0	0
p1i30	1	1	0
p2i1	0	0	0
p2i2	0	0	0
p2i3	0	0	1
p2i4	0	0	0
p2i5	0	0	0
p2i6	0	0	0
p2i7	0	0	0
p2i8	0	0	0

Item	Presence of Subgoals	Number of Subgoals	Relative Definition of Variables
p2i9	0	0	0
p2i10	0	0	0
p2i11	1	1	0
p2i12	0	0	0
p2i13	0	0	0
p2i14	0	0	0
p2i15	0	0	0
p2i16	0	0	0
p2i17	0	0	0
p2i18	1	1	0
p2i19	0	0	0
p2i20	0	0	0
p2i21	0	0	0
p2i22	0	0	0
p2i23	0	0	0
p2i24	0	0	0
p2i25	0	0	0
p2i26	0	0	0
p2i27	1	1	0
p3i1	0	0	0
p3i2	0	0	0
p3i3	0	0	0
p3i4	0	0	0
p3i5	0	0	0
p3i6	0	0	0
p3i7	0	0	0
p3i8	0	0	0
p3i9	0	0	0
p3i10	0	0	0
p3i11	0	0	0
p3i12	0	0	0
p3i13	0	0	0
p3i14	0	0	0
p3i15	0	0	0
p3i16	0	0	0
p3i17	0	0	0
p3i18	0	0	0
p3i19	0	0	0
p3i20	0	0	0

Item	Presence of Subgoals	Number of Subgoals	Relative Definition of Variables
p3i21	0	0	1
p3i22	0	0	0
p3i23	0	0	0
p3i24	0	0	0
p3i25	0	0	0
p3i26	0	0	0
p3i27	0	0	0
p3i28	0	0	0
p3i29	0	0	0

Table B 12. *Cognitive variable scores for number knowledge.*

Item	Number Knowledge	Single-digit	Double-digit	Triple-digit	Four-digit+	Fraction/ Decimal
p1i1	1	1	0	0	0	0
p1i2	5	1	1	0	0	1
p1i3	1	1	0	0	0	0
p1i5	1	1	0	0	0	0
p1i6	5	1	1	0	0	1
p1i7	0	0	0	0	0	0
p1i8	1	1	0	0	0	0
p1i9	1	1	0	0	0	0
p1i10	3	0	0	1	0	0
p1i11	5	0	1	0	0	1
p1i12	1	1	0	0	0	0
p1i13	2	1	1	0	0	0
p1i14	2	1	1	0	0	0
p1i15	2	1	1	0	0	0
p1i16	5	1	1	0	0	1
p1i17	4	0	0	1	1	0
p1i18	1	1	0	0	0	0
p1i19	2	1	1	0	0	0
p1i20	2	0	1	0	0	0
p1i21	1	1	0	0	0	0
p1i22	2	1	1	0	0	0
p1i23	2	1	1	0	0	0
p1i24	4	0	0	0	1	0
p1i25	2	0	1	0	0	0
p1i26	2	1	1	0	0	0
p1i27	3	1	1	1	0	0
p1i28	3	0	1	1	0	0
p1i29	2	1	1	0	0	0
p1i30	2	1	1	0	0	0
p2i1	2	1	1	0	0	0
p2i2	5	1	0	0	0	1
p2i3	4	1	0	0	1	0
p2i4	1	1	0	0	0	0
p2i5	5	0	1	0	0	1
p2i6	1	1	0	0	0	0
p2i7	5	0	0	1	0	1

Item	Number Knowledge	Single-digit	Double-digit	Triple-digit	Four-digit+	Fraction/ Decimal
p2i8	5	0	1	0	0	1
p2i9	5	0	0	1	1	1
p2i10	5	1	1	1	0	1
p2i11	5	0	1	0	0	1
p2i12	5	0	1	0	0	1
p2i13	5	1	0	0	1	1
p2i14	5	1	0	0	0	1
p2i15	4	0	1	0	1	0
p2i16	5	0	1	1	0	1
p2i17	2	1	1	0	0	0
p2i18	5	0	1	0	0	1
p2i19	5	1	1	0	0	1
p2i20	5	1	1	0	0	1
p2i21	2	1	1	0	0	0
p2i22	2	0	1	0	0	0
p2i23	2	1	1	0	0	0
p2i24	5	0	1	0	0	1
p2i25	5	0	0	0	0	1
p2i26	5	1	0	0	0	1
p2i27	5	1	0	0	0	1
p3i1	2	0	1	0	0	0
p3i2	5	1	0	0	0	1
p3i3	5	0	1	0	0	1
p3i4	3	0	1	1	0	0
p3i5	5	1	1	0	0	1
p3i6	2	1	1	0	0	0
p3i7	5	1	1	0	0	1
p3i8	2	1	1	0	0	0
p3i9	4	0	1	0	1	0
p3i10	1	1	0	0	0	0
p3i11	2	1	1	0	0	0
p3i12	1	1	0	0	0	0
p3i13	2	1	1	0	0	0
p3i14	5	0	0	0	0	1
p3i15	0	0	0	0	0	0
p3i16	5	1	0	0	0	1
p3i17	5	1	0	0	0	1
p3i18	1	1	0	0	0	0

Item	Number Knowledge	Single-digit	Double-digit	Triple-digit	Four-digit+	Fraction/ Decimal
p3i19	0	0	0	0	0	0
p3i20	1	1	0	0	0	0
p3i21	1	1	0	0	0	0
p3i22	5	1	0	0	0	1
p3i23	1	1	0	0	0	0
p3i24	5	1	0	0	0	1
p3i25	1	1	0	0	0	0
p3i26	5	0	0	0	0	1
p3i27	4	0	1	1	1	0
p3i28	2	1	1	0	0	0
p3i29	5	1	1	0	0	1

Table B 13. *Cognitive variable scores for alternate procedural knowledge and semantic memory.*

Item	Alt. Procedural Knowledge	1. Multiple Steps	2. Algebraic Equations	3. Mixed fractions	Semantic Memory
p1i1	0	0	0	0	2
p1i2	0	0	0	0	2
p1i3	0	0	0	0	2
p1i5	0	0	0	0	2
p1i6	0	0	0	0	1
p1i7	0	0	0	0	2
p1i8	2	1	1	0	1
p1i9	0	0	0	0	1
p1i10	0	0	0	0	1
p1i11	3	0	0	1	1
p1i12	0	0	0	0	1
p1i13	1	1	0	0	3
p1i14	0	0	0	0	2
p1i15	1	1	0	0	3
p1i16	1	1	0	0	3
p1i17	0	0	0	0	3
p1i18	1	1	0	0	3
p1i19	2	1	1	0	1
p1i20	2	1	1	0	1
p1i21	2	1	1	0	1
p1i22	2	1	1	0	1
p1i23	2	1	1	0	1
p1i24	1	1	0	0	1
p1i25	2	0	1	0	1
p1i26	2	0	1	0	1
p1i27	2	1	1	0	1
p1i28	2	1	1	0	1
p1i29	2	0	1	0	1
p1i30	2	1	1	0	1
p2i1	2	1	1	0	1
p2i2	2	1	1	0	1
p2i3	2	0	1	0	1
p2i4	3	1	1	1	1
p2i5	2	1	1	0	1

Item	Alt. Procedural Knowledge	1. Multiple Steps	2. Algebraic Equations	3. Mixed fractions	Semantic Memory
p2i6	0	0	0	0	1
p2i7	0	0	0	0	1
p2i8	0	0	0	0	1
p2i9	0	0	0	0	1
p2i10	0	0	0	0	1
p2i11	1	1	0	0	1
p2i12	1	1	0	0	1
p2i13	1	1	0	0	1
p2i14	3	1	0	1	1
p2i15	1	1	0	0	1
p2i16	1	1	0	0	1
p2i17	1	1	0	0	2
p2i18	1	1	0	0	2
p2i19	0	0	0	0	2
p2i20	1	1	0	0	2
p2i21	2	0	1	0	2
p2i22	1	1	0	0	2
p2i23	0	0	0	0	1
p2i24	0	0	0	0	1
p2i25	0	0	0	0	1
p2i26	0	0	0	0	1
p2i27	1	1	0	0	1
p3i1	2	0	1	0	4
p3i2	2	0	1	0	4
p3i3	2	0	1	0	4
p3i4	2	0	1	0	4
p3i5	2	0	1	0	4
p3i6	2	0	1	0	4
p3i7	2	0	1	0	4
p3i8	0	0	0	0	2
p3i9	0	0	0	0	2
p3i10	0	0	0	0	2
p3i11	1	1	0	0	2
p3i12	0	0	0	0	2
p3i13	1	1	0	0	2
p3i14	0	0	0	0	1
p3i15	0	0	0	0	1
p3i16	0	0	0	0	1

Item	Alt. Procedural Knowledge	1. Multiple Steps	2. Algebraic Equations	3. Mixed fractions	Semantic Memory
p3i17	0	0	0	0	1
p3i18	1	1	0	0	1
p3i19	0	0	0	0	1
p3i20	0	0	0	0	1
p3i21	2	0	1	0	1
p3i22	1	1	0	0	1
p3i23	1	1	0	0	1
p3i24	1	1	0	0	1
p3i25	0	0	0	0	1
p3i26	2	0	1	0	1
p3i27	2	0	1	0	1
p3i28	0	0	0	0	1
p3i29	2	1	1	0	1

Table B 14. *Cognitive variable scores for procedural knowledge.*

Item	Procedural Knowledge	1. Integers	2. Fractions	3. Proportions	4. Decimals	5. Negative Numbers	6. Square Roots	Number of Procedures
p1i1	0	0	0	0	0	0	0	0
p1i2	2	0	1	0	0	0	0	2
p1i3	0	0	0	0	0	0	0	0
p1i5	0	0	0	0	0	0	0	0
p1i6	2	0	1	0	0	0	0	1
p1i7	0	0	0	0	0	0	0	0
p1i8	1	1	0	0	0	0	0	1
p1i9	0	0	0	0	0	0	0	0
p1i10	4	0	0	0	1	0	0	1
p1i11	2	1	1	0	0	0	0	2
p1i12	0	0	0	0	0	0	0	0
p1i13	1	1	0	0	0	0	0	1
p1i14	0	0	0	0	0	0	0	0
p1i15	5	1	0	0	0	1	0	2
p1i16	4	1	0	0	1	0	0	2
p1i17	0	0	0	0	0	0	0	0
p1i18	1	1	0	0	0	0	0	1
p1i19	6	1	0	0	0	0	1	2
p1i20	6	1	0	0	0	0	1	2
p1i21	6	1	0	0	0	0	1	2
p1i22	6	1	0	0	0	0	1	2
p1i23	6	1	0	0	0	0	1	2
p1i24	1	1	0	0	0	0	0	1
p1i25	1	1	0	0	0	0	0	1

Item	Procedural Knowledge	1. Integers	2. Fractions	3. Proportions	4. Decimals	5. Negative Numbers	6. Square Roots	Number of Procedures
p1i26	0	0	0	0	0	0	0	0
p1i27	1	1	0	0	0	0	0	1
p1i28	1	1	0	0	0	0	0	1
p1i29	5	1	0	0	0	1	0	2
p1i30	1	1	0	0	0	0	0	1
p2i1	1	1	0	0	0	0	0	1
p2i2	5	1	1	0	0	1	0	3
p2i3	0	0	0	0	0	0	0	0
p2i4	1	1	0	0	0	0	0	1
p2i5	4	1	0	0	1	0	0	2
p2i6	0	0	0	0	0	0	0	0
p2i7	2	0	1	0	0	0	0	1
p2i8	0	0	0	0	0	0	0	1
p2i9	2	0	1	0	0	0	0	1
p2i10	2	0	1	0	0	0	0	1
p2i11	2	0	1	0	0	0	0	1
p2i12	2	0	1	0	0	0	0	1
p2i13	2	0	1	0	0	0	0	1
p2i14	2	0	1	1	0	0	0	2
p2i15	2	0	1	1	0	0	0	1
p2i16	4	0	1	1	1	0	0	2
p2i17	2	0	1	1	0	0	0	1
p2i18	4	0	1	1	1	0	0	3
p2i19	4	0	1	1	1	0	0	3
p2i20	6	0	0	0	1	0	1	2
p2i21	0	0	0	0	0	0	0	0

Item	Procedural Knowledge	1. Integers	2. Fractions	3. Proportions	4. Decimals	5. Negative Numbers	6. Square Roots	Number of Procedures
p2i22	0	0	0	0	0	0	0	0
p2i23	4	0	0	0	1	0	0	1
p2i24	4	0	0	1	1	0	0	2
p2i25	2	0	1	0	0	0	0	2
p2i26	2	0	1	0	0	0	0	1
p2i27	2	0	1	0	0	0	0	1
p3i1	0	0	0	0	0	0	0	0
p3i2	0	0	0	0	0	0	0	0
p3i3	0	0	0	0	0	0	0	0
p3i4	0	0	0	0	0	0	0	0
p3i5	0	0	0	0	0	0	0	0
p3i6	0	0	0	0	0	0	0	0
p3i7	0	0	0	0	0	0	0	0
p3i8	0	0	0	0	0	0	0	0
p3i9	5	0	0	0	0	1	0	1
p3i10	0	0	0	0	0	0	0	0
p3i11	6	0	0	0	0	1	1	2
p3i12	5	0	0	0	0	1	0	1
p3i13	0	0	0	0	0	0	0	0
p3i14	5	0	1	0	0	1	0	0
p3i15	0	0	0	0	0	0	0	0
p3i16	0	0	0	0	0	0	0	0
p3i17	0	0	0	0	0	0	0	0
p3i18	5	0	0	0	0	1	0	1
p3i19	0	0	0	0	0	0	0	0
p3i20	0	0	0	0	0	0	0	0

Item	Procedural Knowledge	1. Integers	2. Fractions	3. Proportions	4. Decimals	5. Negative Numbers	6. Square Roots	Number of Procedures
p3i21	0	0	0	0	0	0	0	0
p3i22	5	0	1	0	0	1	0	2
p3i23	5	0	0	0	0	1	0	1
p3i24	2	0	1	0	0	0	0	1
p3i25	4	0	0	0	0	1	0	1
p3i26	0	0	0	0	0	0	0	0
p3i27	0	0	0	0	0	0	0	0
p3i28	0	0	0	0	0	0	0	0
p3i29	2	0	1	0	0	0	0	0

Table B 15. *Cognitive variable scores for number of computations, operands, and meta-cognition processes.*

Item	Number of Computations	Number of Operands	Meta-Cognition Process
p1i1	0	1	0
p1i2	1	1	0
p1i3	0	0	1
p1i5	0	0	0
p1i6	1	0	0
p1i7	0	0	1
p1i8	3	2	0
p1i9	1	1	1
p1i10	0	1	1
p1i11	3	1	0
p1i12	0	0	1
p1i13	10	2	1
p1i14	2	0	0
p1i15	4	2	0
p1i16	2	2	0
p1i17	0	0	0
p1i18	11	2	1
p1i19	9	3	0
p1i20	4	3	0
p1i21	4	3	0
p1i22	16	2	0
p1i23	4	3	0
p1i24	3	2	1
p1i25	1	1	0
p1i26	4	2	0
p1i27	1	2	1
p1i28	3	3	1
p1i29	4	1	0
p1i30	1	2	1
p2i1	2	2	0
p2i2	2	2	0
p2i3	1	1	1
p2i4	2	2	0
p2i5	2	2	0

Item	Number of Computations	Number of Operands	Meta-Cognition Process
p2i6	0	0	1
p2i7	1	1	0
p2i8	3	2	0
p2i9	2	1	0
p2i10	2	1	0
p2i11	2	1	1
p2i12	2	2	0
p2i13	3	2	0
p2i14	5	2	0
p2i15	3	2	1
p2i16	3	2	0
p2i17	3	2	1
p2i18	2	2	1
p2i19	1	1	0
p2i20	2	2	0
p2i21	0	1	1
p2i22	3	3	0
p2i23	2	1	0
p2i24	3	1	0
p2i25	1	1	0
p2i26	1	1	0
p2i27	2	2	1
p3i1	0	2	1
p3i2	0	2	1
p3i3	0	2	1
p3i4	0	2	1
p3i5	0	2	1
p3i6	0	1	1
p3i7	0	2	1
p3i8	0	1	1
p3i9	1	1	0
p3i10	0	0	0
p3i11	7	5	1
p3i12	1	1	0
p3i13	5	3	1
p3i14	4	0	1
p3i15	0	0	1

Item	Number of Computations	Number of Operands	Meta-Cognition Process
p3i16	0	0	1
p3i17	0	0	1
p3i18	1	1	0
p3i19	0	0	1
p3i20	0	0	1
p3i21	0	0	0
p3i22	2	1	0
p3i23	1	1	0
p3i24	2	1	0
p3i25	0	0	1
p3i26	0	1	0
p3i27	0	3	1
p3i28	0	0	1
p3i29	0	3	1

Table B 16. *Cognitive variable scores for decision processing component.*

Item	Decisions Processing Confirmation	Bottom-Up Processing	Top-Down Processing	Functional Distractors
p1i1	1	0	1	0
p1i2	0	0	0	1
p1i3	0	0	0	2
p1i5	1	0	1	1
p1i6	1	0	1	1
p1i7	0	0	0	0
p1i8	0	0	0	1
p1i9	1	1	0	3
p1i10	1	0	1	3
p1i11	0	0	0	1
p1i12	1	1	0	3
p1i13	1	1	0	3
p1i14	0	0	0	1
p1i15	0	0	0	3
p1i16	0	0	0	3
p1i17	1	0	1	1
p1i18	1	0	1	2
p1i19	1	1	0	3
p1i20	0	0	0	3
p1i21	0	0	0	3
p1i22	1	1	0	3
p1i23	0	0	0	2
p1i24	1	1	0	2
p1i25	1	0	1	3
p1i26	1	1	0	3
p1i27	1	0	1	3
p1i28	1	0	1	3
p1i29	1	1	0	2
p1i30	1	0	1	3
p2i1	0	0	0	1
p2i2	0	0	0	2
p2i3	1	0	1	3
p2i4	0	0	0	0
p2i5	0	0	0	3
p2i6	1	1	0	2

Item	Decisions Processing Confirmation	Bottom-Up Processing	Top-Down Processing	Functional Distractors
p2i7	0	0	0	0
p2i8	1	0	1	3
p2i9	0	0	0	2
p2i10	0	0	0	1
p2i11	0	0	0	1
p2i12	0	0	0	1
p2i13	0	0	0	1
p2i14	0	0	0	2
p2i15	0	0	0	3
p2i16	0	0	0	2
p2i17	0	0	0	2
p2i18	0	0	0	3
p2i19	0	0	0	1
p2i20	0	0	0	3
p2i21	1	0	1	0
p2i22	0	0	0	3
p2i23	0	0	0	3
p2i24	0	0	0	3
p2i25	0	0	0	3
p2i26	0	0	0	2
p2i27	0	0	0	3
p3i1	1	0	1	0
p3i2	1	0	1	0
p3i3	1	0	1	1
p3i4	1	0	1	3
p3i5	1	0	1	2
p3i6	1	0	1	1
p3i7	1	0	1	2
p3i8	1	0	1	1
p3i9	0	0	0	2
p3i10	1	1	0	2
p3i11	0	0	0	3
p3i12	0	0	0	2
p3i13	0	0	0	3
p3i14	1	1	0	3
p3i15	1	0	1	3
p3i16	1	1	0	1

Item	Decisions Processing Confirmation	Bottom-Up Processing	Top-Down Processing	Functional Distractors
p3i17	1	1	0	3
p3i18	0	0	0	0
p3i19	1	1	0	3
p3i20	1	1	0	1
p3i21	1	1	0	3
p3i22	0	0	0	3
p3i23	0	0	0	2
p3i24	0	0	0	2
p3i25	1	1	0	3
p3i26	1	0	1	1
p3i27	1	0	1	3
p3i28	1	1	0	3
p3i29	1	0	1	3

Table B 17. *Zero-order correlations between cognitive variables and IRT psychometric properties.*

	1PL	2PL	
	b	**a**	**b**
Translation			
Total Encoding	0.170	-0.036	0.187
Mathematical	0.071	0.050	0.132
Contextual	0.179	-0.087	0.157
Stem	0.154	0.085	0.194
Content Words	0.103	0.077	0.145
Text Comprehensions			
Flesch Reading Ease Test	-0.075	0.028	-0.067
Flesch-Kincaid Grade Level Test	0.110	0.018	0.111
Text Comprehension – LSA (Average)	-0.053	-0.097	-0.107
Mathematical Propositions			
Assignment Propositions	-0.248	0.122	-0.233
Relation Propositions	0.082	-0.121	0.061
Contextual Propositions			
Total Number of Propositions	0.186	0.047	0.211
Total Proposition Density	0.163	-0.013	0.160
Number of Predicate Propositions	0.144	-0.031	0.164
Predicate Density	0.095	-0.149	0.089
Number of Modifier Propositions	0.183	0.109	0.200
Modifier Density	0.157	0.072	0.145
Number of Connective Propositions	0.082	-0.019	0.121
Connector Density	-0.003	0.018	0.028
Total Number of Unique Arguments	0.203	0.029	0.245
Unique Argument Density	0.119	-0.095	0.115
Total Number of Arguments	0.200	0.019	0.226
Total Argument Ratio	0.212	-0.014	0.225
Max. Number of Arguments	0.207	0.049	0.232
Relevant Propositions	0.227	0.019	0.256
Density of Relevant Propositions	0.026	-0.081	0.017
Relevant Words	0.171	0.079	0.219
Density of Relevant Words	-0.003	-0.081	-0.011
Irrelevant Propositions	-0.008	0.053	-0.006
Density of Irrelevant Propositions	-0.026	0.081	-0.017
Irrelevant Words	0.017	0.038	0.017
Density of Irrelevant Words	0.003	0.081	0.011

	1PL	2PL	
	b	**a**	**b**
Encode Diagram	0.165	0.114	0.193
Integration			
Translate Word Equation	0.020	-0.104	0.035
Given Equation – in Stem	-0.079	0.128	-0.083
Generate Eq. or Possible Values	0.100	0.014	0.139
Access Equation	0.013	0.107	0.023
Auxiliary Diagram	0.047	0.230	0.083
Translate Diagram	0.157	-0.028	0.163
Visualization	0.135	0.141	0.135
Semantic Memory	-0.204	-0.050	-0.234
Solution Planning			
Presence of Subgoals	0.189	-0.120	0.167
Number of Subgoals	0.178	-0.143	0.156
Relative Definition of Variables	0.109	-0.104	0.092
Solution Execution			
Number Knowledge	0.142	-0.066	0.138
1. Single-digit	0.000	-0.020	-0.014
2. Double-digit	0.126	-0.005	0.129
3. Triple-digit	-0.046	0.067	-0.016
4. Four-digit +	-0.008	-0.036	-0.017
5. Fraction/Decimal	0.070	-0.018	0.080
Alt. Procedural Knowledge	-0.029	0.216	0.025
1. Multiple Steps	0.098	0.167	0.143
2. Algebraic Equations	-0.048	0.151	0.003
3. Mixed fractions	-0.120	0.083	-0.105
Procedural Knowledge	0.148	0.203	0.183
1. Integers	-0.096	0.144	-0.084
2. Fractions	0.090	0.081	0.143
3. Proportions	0.172	0.033	0.190
4. Decimals	0.237	-0.011	0.238
5. Negative Numbers	-0.063	0.039	-0.059
6. Square Roots	0.124	0.186	0.124
Number of Procedures	0.120	0.171	0.142
Number of Computations	0.119	0.196	0.140
Number of Operands	0.126	0.108	0.169
Meta-Cognition Process	0.115	-0.347	0.054
Decision Processing			
Decision Processing Confirmation	0.056	-0.178	0.025

	1PL	2PL	
	b	**a**	**b**
Bottom-Up Processing	0.155	-0.174	0.094
Top-Down Processing	-0.078	-0.039	-0.057
Functional Distractors	0.747	-0.340	0.676

APPENDIX C
MODEL ESTIMATES

Table C 1. *Estimates from the LLTM null model.*

Parameter	Estimate	Standard Error	t-value
μ	0.9989	0.02797	35.71*
Intercept	-1.0583	0.01845	-57.38*

*$p < 0.05$

Table C 2. *Estimates from the LLTM saturated model.*

Parameter	Estimate	Standard Error	t-value
i1	-3.6625	0.0958	-38.23*
i2	-2.3673	0.0623	-38*
i3	-0.3596	0.0467	-7.69*
i5	-1.8042	0.0548	-32.93*
i6	-0.6525	0.0474	-13.77*
i7	-3.3421	0.0848	-39.41*
i8	-2.4125	0.0630	-38.27*
i9	-0.3334	0.0467	-7.14*
i10	-0.3789	0.0468	-8.1*
i11	-2.0307	0.0574	-35.36*
i12	0.5888	0.0475	12.39*
i13	-0.3282	0.0467	-7.03*
i14	-3.0697	0.0770	-39.86*
i15	-1.8196	0.0550	-33.11*
i16	-0.9303	0.0484	-19.22*
i17	-1.7611	0.0543	-32.41*
i18	-2.0827	0.0581	-35.85*
i19	-0.7271	0.0476	-15.27*
i20	-1.1980	0.0498	-24.06*
i21	-1.1453	0.0495	-23.14*
i22	-1.0286	0.0489	-21.05*
i23	-1.6175	0.0530	-30.54*
i24	-1.8247	0.0550	-33.17*
i25	-0.8924	0.0482	-18.5*
i26	0.3362	0.0469	7.17*
i27	-1.4132	0.0513	-27.57*
i28	-0.4034	0.0468	-8.62*
i29	-1.3958	0.0511	-27.3*
i30	-0.7656	0.0477	-16.04*
i31	-2.4160	0.0631	-38.29*
i32	-1.4931	0.0519	-28.77*
i33	-0.4228	0.0468	-9.03*
i34	-3.1030	0.0779	-39.84*
i35	-1.4220	0.0513	-27.7*
i36	-1.4796	0.0518	-28.58*
i37	-3.0320	0.0760	-39.88*

i38	-0.2099	0.0466	-4.51*
i39	-2.1666	0.0593	-36.57*
i40	-2.0024	0.0571	-35.09*
i41	-0.8045	0.0479	-16.8*
i42	-2.2448	0.0604	-37.18*
i43	-2.0974	0.0583	-35.98*
i44	-0.7161	0.0476	-15.05*
i45	-0.6833	0.0475	-14.4*
i46	-1.6652	0.0534	-31.18*
i47	-0.3719	0.0468	-7.95*
i48	0.0065	0.0465	0.14
i49	-2.0307	0.0574	-35.36*
i50	0.8829	0.0487	18.14*
i51	-3.3625	0.0854	-39.36*
i52	-0.5058	0.0470	-10.76*
i53	-0.1094	0.0465	-2.35*
i54	0.9198	0.0488	18.83*
i55	0.2483	0.0468	5.31*
i56	-0.2359	0.0466	-5.06*
i57	0.4588	0.0472	9.73*
i58	-3.1603	0.0795	-39.78*
i59	-2.6553	0.0675	-39.34*
i60	-2.4956	0.0645	-38.7*
i61	-1.2308	0.0500	-24.61*
i62	-2.0137	0.0572	-35.2*
i63	0.8292	0.0484	17.12*
i64	-2.0052	0.0571	-35.11*
i65	-3.0108	0.0755	-39.88*
i66	-1.4154	0.0513	-27.6*
i67	-0.3317	0.0467	-7.1*
i68	-0.9113	0.0483	-18.86*
i69	-1.0442	0.0489	-21.34*
i70	-0.7822	0.0478	-16.37*
i71	-0.9571	0.0485	-19.73*
i72	-0.7878	0.0478	-16.47*
i73	-1.5225	0.0521	-29.21*
i74	-1.0014	0.0487	-20.55*
i75	-2.8760	0.0722	-39.81*
i76	-1.5431	0.0523	-29.5*
i77	-2.4373	0.0635	-38.41*

i78	-0.5913	0.0472	-12.53*
i79	-0.7307	0.0476	-15.35*
i80	-2.1122	0.0585	-36.11*
i81	-1.1193	0.0493	-22.68*
i82	-1.1473	0.0495	-23.18*
i83	-1.8612	0.0554	-33.59*
i84	0.0117	0.0465	0.25
i85	-1.2911	0.0504	-25.62*
i86	1.1712	0.0502	23.32*
μ	1.4181	0.0387	36.62*

*$p < 0.05$

Table C 3. *Estimated item difficulties for items based on LLTM stimulus feature weights.*

reference	item	$\acute{b}$
p1i1	1	-1.2516
p1i2	2	-1.4726
p1i3	3	-1.1465
p1i5	5	-2.3790
p1i6	6	-1.4452
p1i7	7	-2.2255
p1i8	8	-1.8304
p1i9	9	-1.3860
p1i10	10	-0.8760
p1i11	11	-1.8965
p1i12	12	-1.6388
p1i13	13	-0.5766
p1i14	14	-2.5771
p1i15	15	-1.7321
p1i16	16	-1.8755
p1i17	17	-2.0103
p1i18	18	-1.7432
p1i19	19	-0.8538
p1i20	20	-1.0552
p1i21	21	-1.2682
p1i22	22	-1.1246
p1i23	23	-1.3713
p1i24	24	-0.5739
p1i25	25	-1.3391
p1i26	26	-1.1450
p1i27	27	-0.6194
p1i28	28	-0.9882
p1i29	29	-0.8510
p1i30	30	-1.4845
p2i1	31	-1.5417
p2i2	32	-1.6622
p2i3	33	-0.9575
p2i4	34	-2.2354
p2i5	35	-1.5808
p2i6	36	-1.7178
p2i7	37	-1.6918
p2i8	38	-0.7546
p2i9	39	-1.7301

reference	item	$\acute{b}$
p2i10	40	-1.4557
p2i11	41	-0.7028
p2i12	42	-1.0007
p2i13	43	-1.0295
p2i14	44	-1.0472
p2i15	45	-0.7942
p2i16	46	-0.7546
p2i17	47	-1.7232
p2i18	48	-0.9311
p2i19	49	-1.6774
p2i20	50	-0.4719
p2i21	51	-1.5680
p2i22	52	-1.1305
p2i23	53	-1.0908
p2i24	54	-1.1749
p2i25	55	-1.1445
p2i26	56	-1.3950
p2i27	57	-0.5549
p3i1	58	-1.6990
p3i2	59	-1.6853
p3i3	60	-1.6485
p3i4	61	-1.7997
p3i5	62	-1.7980
p3i6	63	-1.1348
p3i7	64	-1.3719
p3i8	65	-2.4335
p3i9	66	-2.1278
p3i10	67	-2.0009
p3i11	68	-1.3945
p3i12	69	-2.2478
p3i13	70	-2.2767
p3i14	71	-1.0157
p3i15	72	-0.5980
p3i16	73	-1.5758
p3i17	74	-1.0983
p3i18	75	-1.2160
p3i19	76	-1.9429
p3i20	77	-1.4242
p3i21	78	-0.5461
p3i22	79	-1.1797

reference	item	$\acute{b}$
p3i23	80	-1.1454
p3i24	81	-1.2463
p3i25	82	-1.0462
p3i26	83	-1.2718
p3i27	84	-0.6191
p3i28	85	-1.6563
p3i29	86	-0.5830

Table C 4. *Estimated item difficulty and item discrimination from 2PL regressions.*

reference	i	$\acute{b}$	$\acute{a}$
p1i1	1	-1.2374	1.1993
p1i2	2	-0.8078	1.2136
p1i3	3	-0.9232	1.2520
p1i5	5	-1.8972	1.3619
p1i6	6	-0.5860	1.3131
p1i7	7	-1.9659	1.1987
p1i8	8	-1.3012	1.3213
p1i9	9	-0.8939	1.3908
p1i10	10	-0.3301	1.0778
p1i11	11	-1.1408	1.1454
p1i12	12	-1.1439	1.3808
p1i13	13	-0.3068	0.9514
p1i14	14	-1.9913	1.3534
p1i15	15	-1.0931	1.5177
p1i16	16	-1.0246	1.3036
p1i17	17	-1.4982	1.2162
p1i18	18	-1.3461	1.3598
p1i19	19	-0.1749	1.3984
p1i20	20	-0.5057	1.3838
p1i21	21	-0.7608	1.6845
p1i22	22	-0.5078	1.5423
p1i23	23	-0.7441	1.6665
p1i24	24	0.1284	1.2221
p1i25	25	-0.7260	1.3701
p1i26	26	-0.5081	1.4138
p1i27	27	0.0465	1.2443
p1i28	28	-0.2947	1.0906
p1i29	29	-0.1276	1.5450
p1i30	30	-0.8136	1.2354
p2i1	31	-1.0083	1.4217
p2i2	32	-0.7215	1.3145
p2i3	33	-0.2772	0.9077
p2i4	34	-1.6751	1.3521
p2i5	35	-0.6654	1.3869
p2i6	36	-1.4148	1.2259
p2i7	37	-0.8300	1.6424
p2i8	38	-0.0800	1.1323
p2i9	39	-0.7972	1.4223

reference	i	$\acute{b}$	$\acute{a}$
p2i10	40	-0.5436	1.3715
p2i11	41	0.0224	1.1993
p2i12	42	-0.1204	1.2672
p2i13	43	-0.1690	1.4726
p2i14	44	-0.1225	1.6149
p2i15	45	-0.1123	1.2126
p2i16	46	0.1624	1.5805
p2i17	47	-1.1162	1.6409
p2i18	48	-0.0757	1.1689
p2i19	49	-0.7323	1.3826
p2i20	50	0.2608	1.6852
p2i21	51	-1.1553	1.2437
p2i22	52	-0.6958	1.4754
p2i23	53	-0.5019	1.6763
p2i24	54	-0.4090	1.2675
p2i25	55	-0.4779	1.2012
p2i26	56	-0.7256	1.2438
p2i27	57	-0.0633	1.0726
p3i1	58	-1.2776	1.2395
p3i2	59	-0.9418	1.2450
p3i3	60	-0.9486	1.1440
p3i4	61	-1.2501	1.2689
p3i5	62	-1.0748	1.2685
p3i6	63	-0.7801	1.1998
p3i7	64	-0.7275	1.2446
p3i8	65	-1.8989	1.2281
p3i9	66	-1.3215	1.3350
p3i10	67	-1.4771	1.3667
p3i11	68	-0.7545	1.2781
p3i12	69	-1.6764	1.4115
p3i13	70	-1.7615	1.2225
p3i14	71	-0.2237	1.1194
p3i15	72	-0.4219	1.1048
p3i16	73	-0.9449	0.9801
p3i17	74	-0.4714	0.9885
p3i18	75	-0.6490	1.5456
p3i19	76	-1.6972	1.1048
p3i20	77	-1.1861	1.1754
p3i21	78	-0.2710	1.1819
p3i22	79	-0.2887	1.2929

reference	i	$\acute{b}$	$\acute{a}$
p3i23	80	-0.6160	1.4978
p3i24	81	-0.4570	1.2202
p3i25	82	-0.6728	1.2150
p3i26	83	-0.4847	1.1454
p3i27	84	0.0733	1.4185
p3i28	85	-1.0316	1.5699
p3i29	86	0.3661	1.3056

APPENDIX D
FIGURES

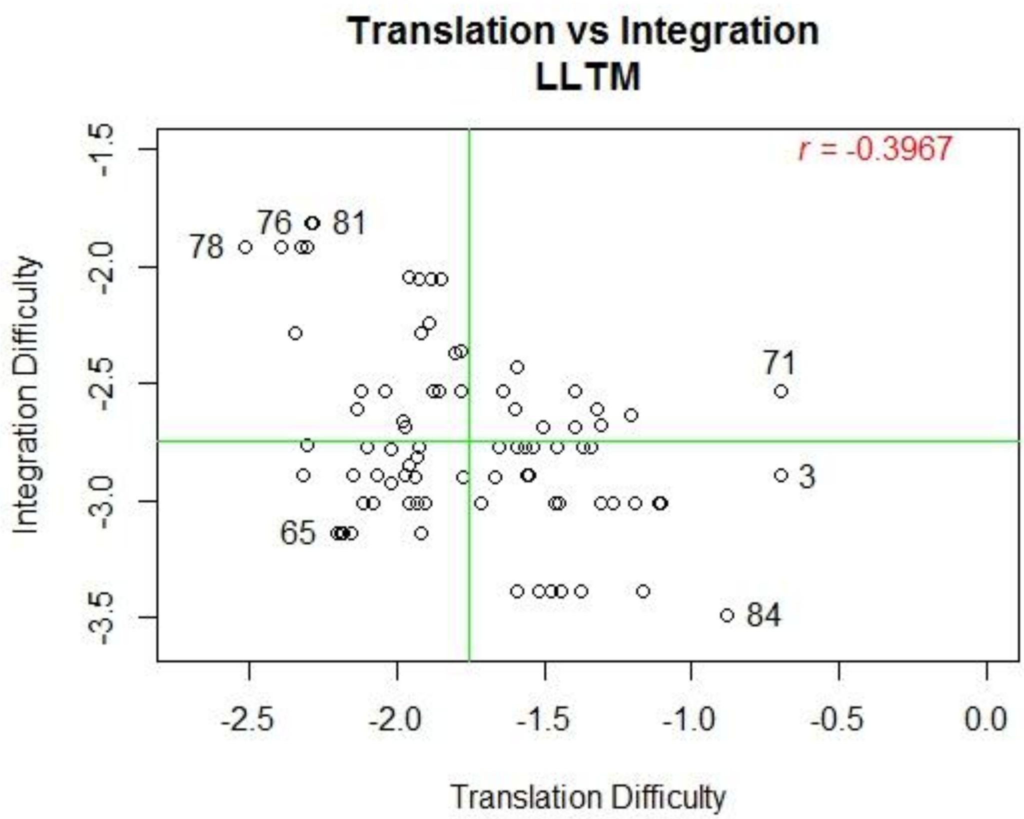


Figure D 1. *Translation component difficulty compared to Integration component difficulty for the LLTM.*

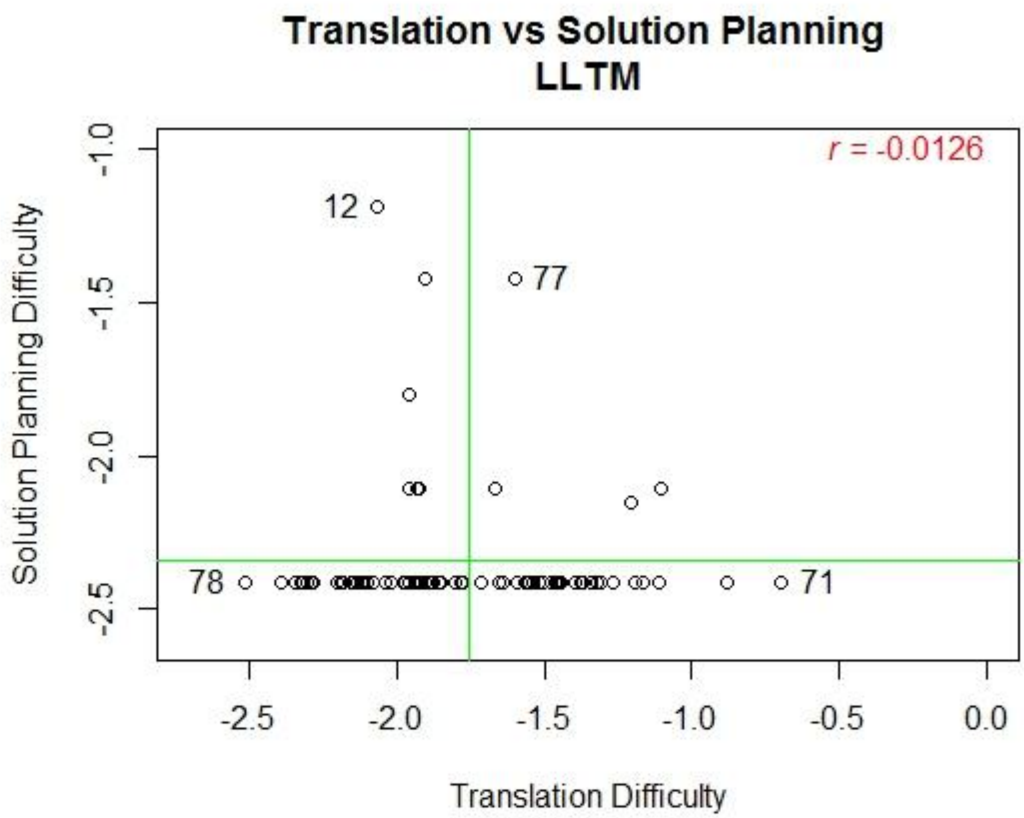


Figure D 2. *Translation component difficulty compared to Solution Planning component difficulty for the LLTM.*

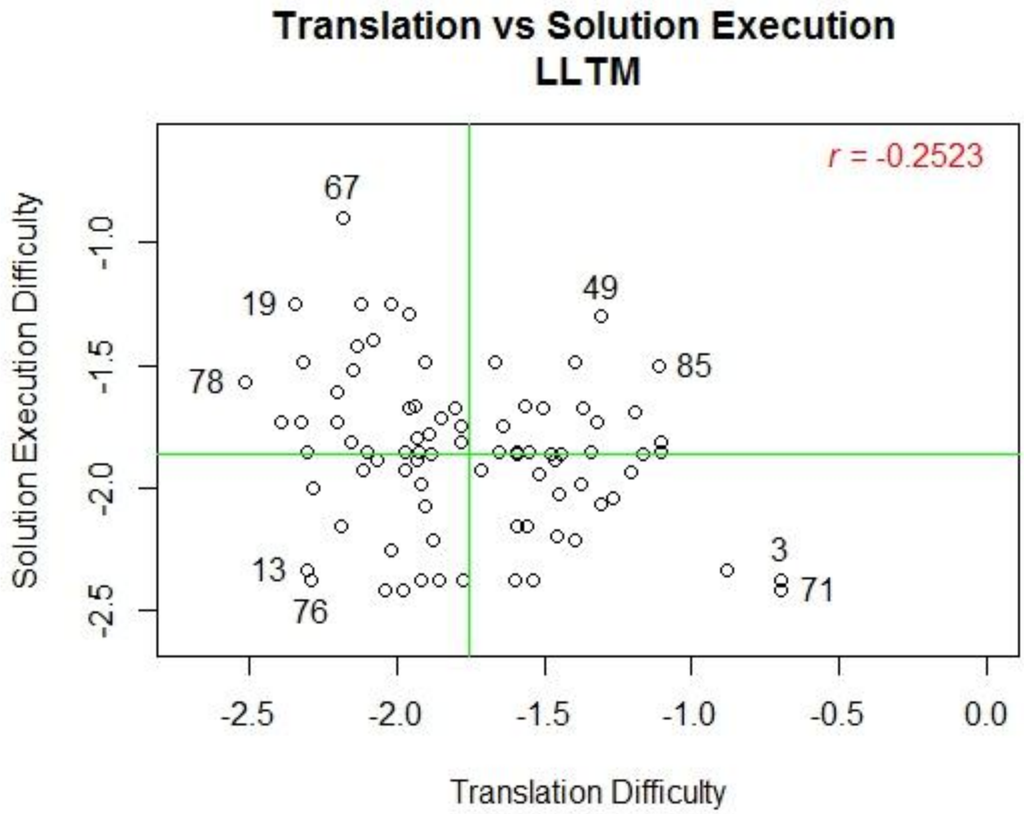


Figure D 3. *Translation component difficulty compared to Solution Execution component difficulty for the LLTM.*

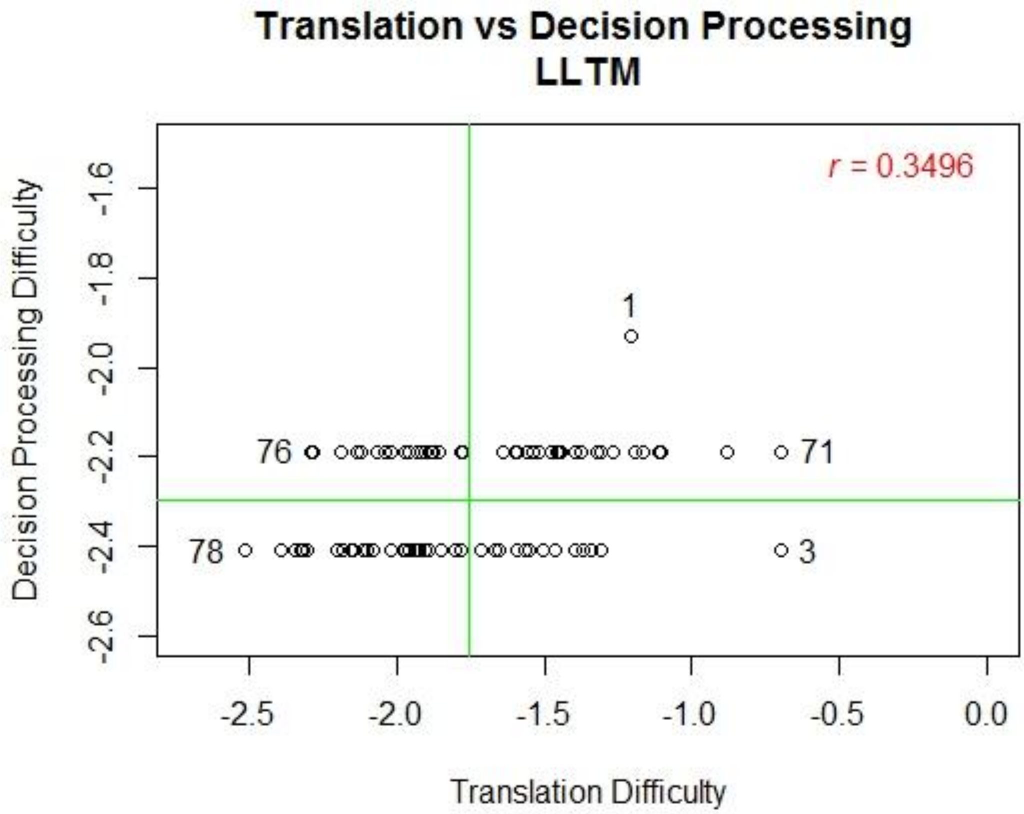


Figure D 4. *Translation component difficulty compared to Decision Processing Component difficulty for the LLTM.*

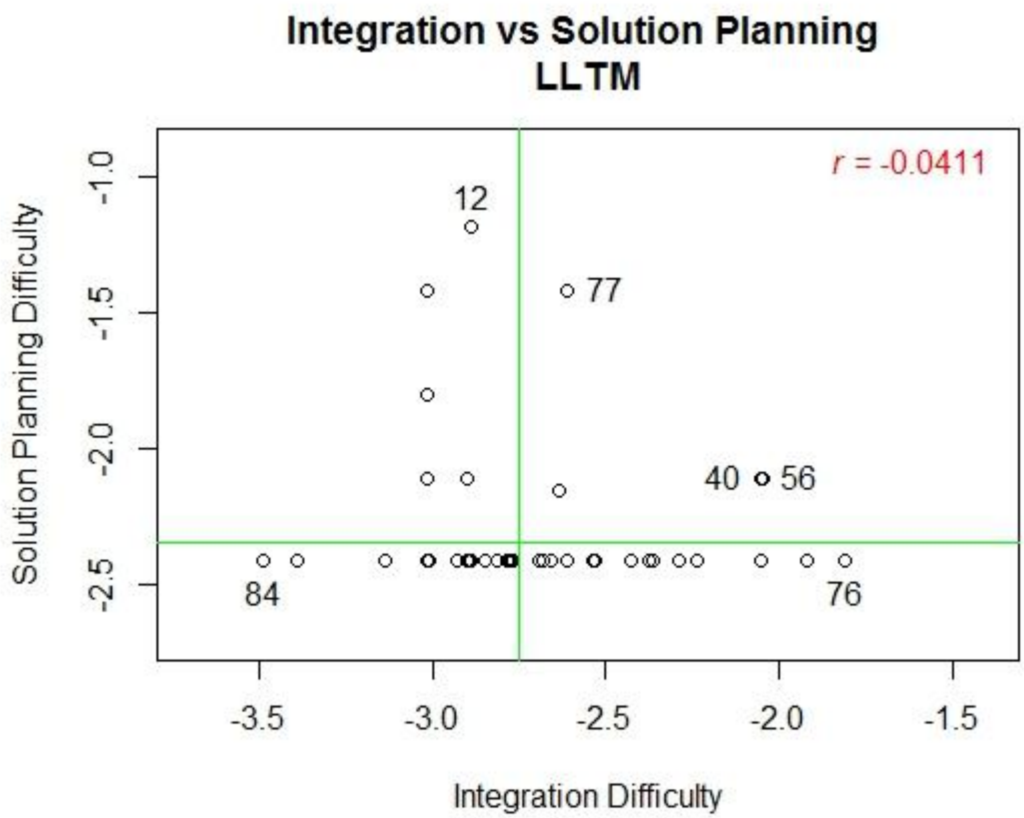


Figure D 5. *Integration component difficulty compared to Solution Planning component difficulty for the LLTM.*

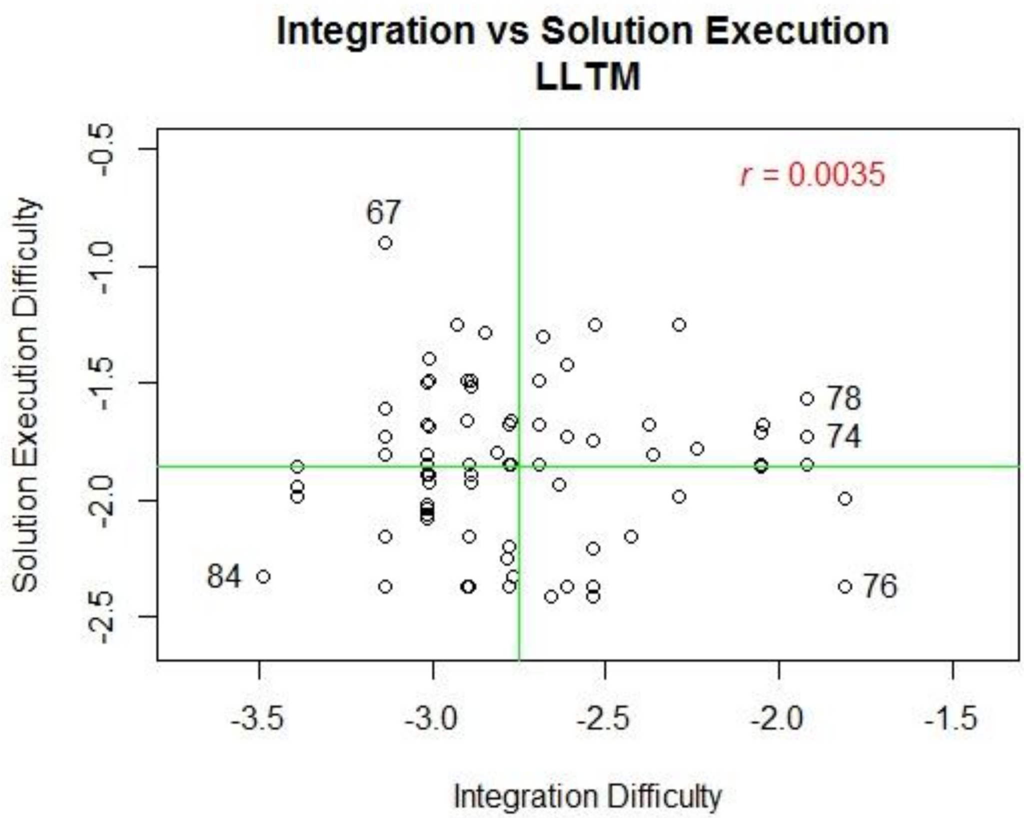


Figure D 6. *Integration component difficulty compared to Solution Execution component difficulty for the LLTM.*

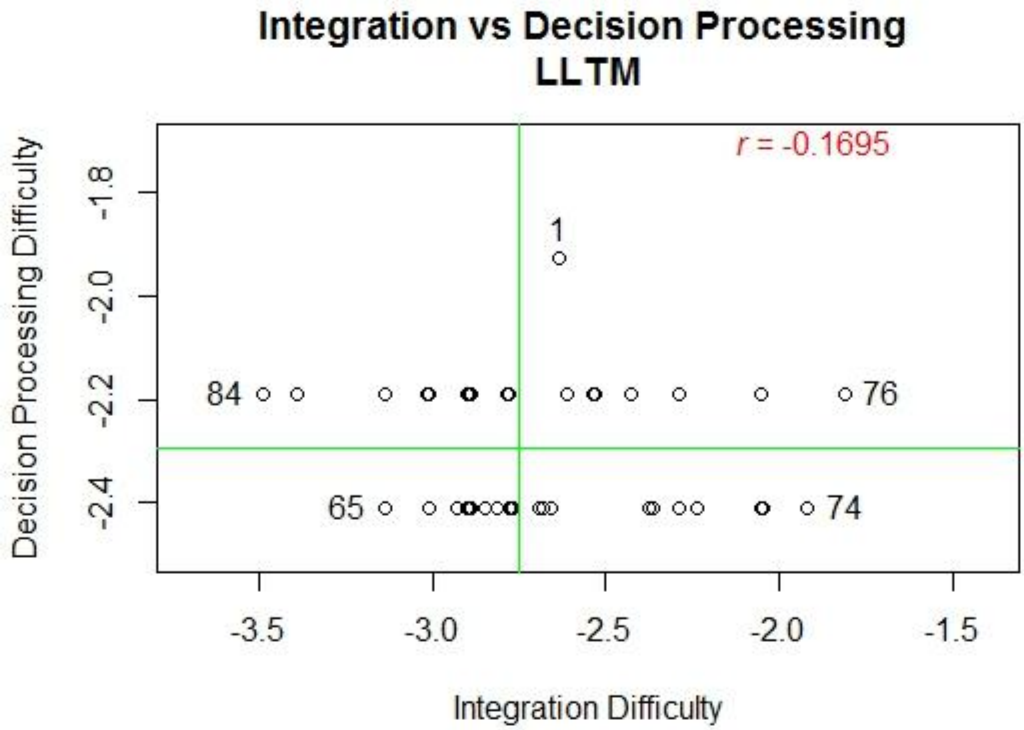


Figure D 7. *Integration component difficulty compared to Decision Processing component difficulty for the LLTM.*

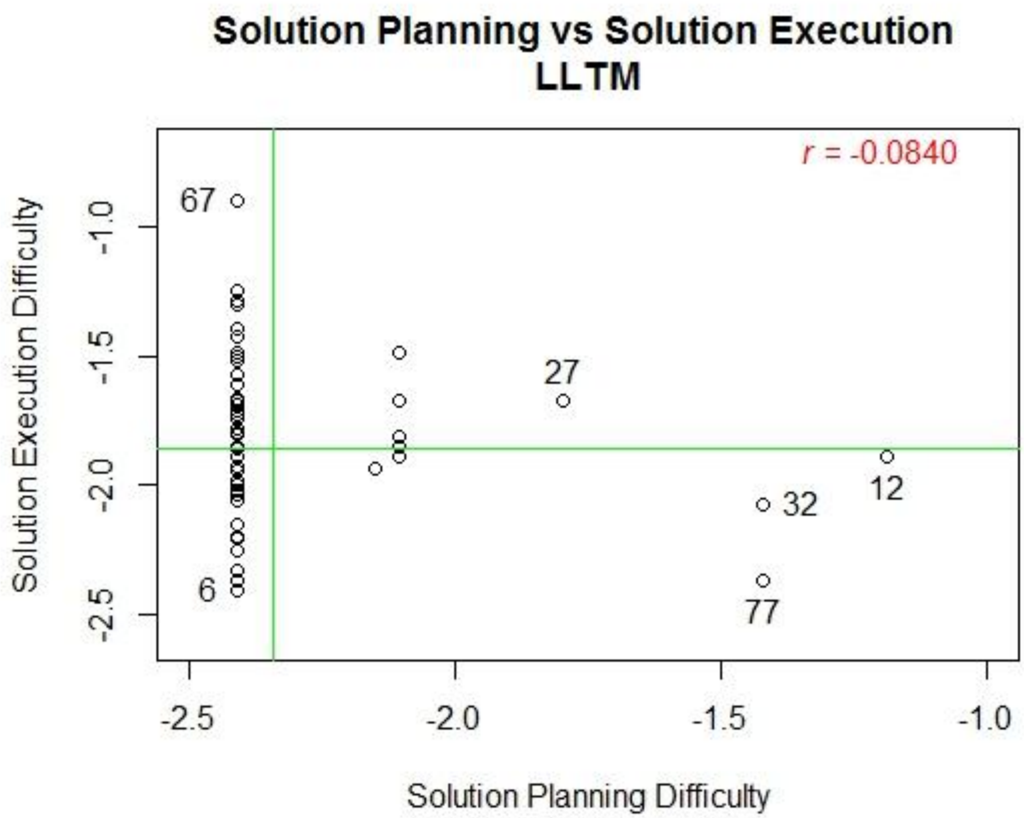
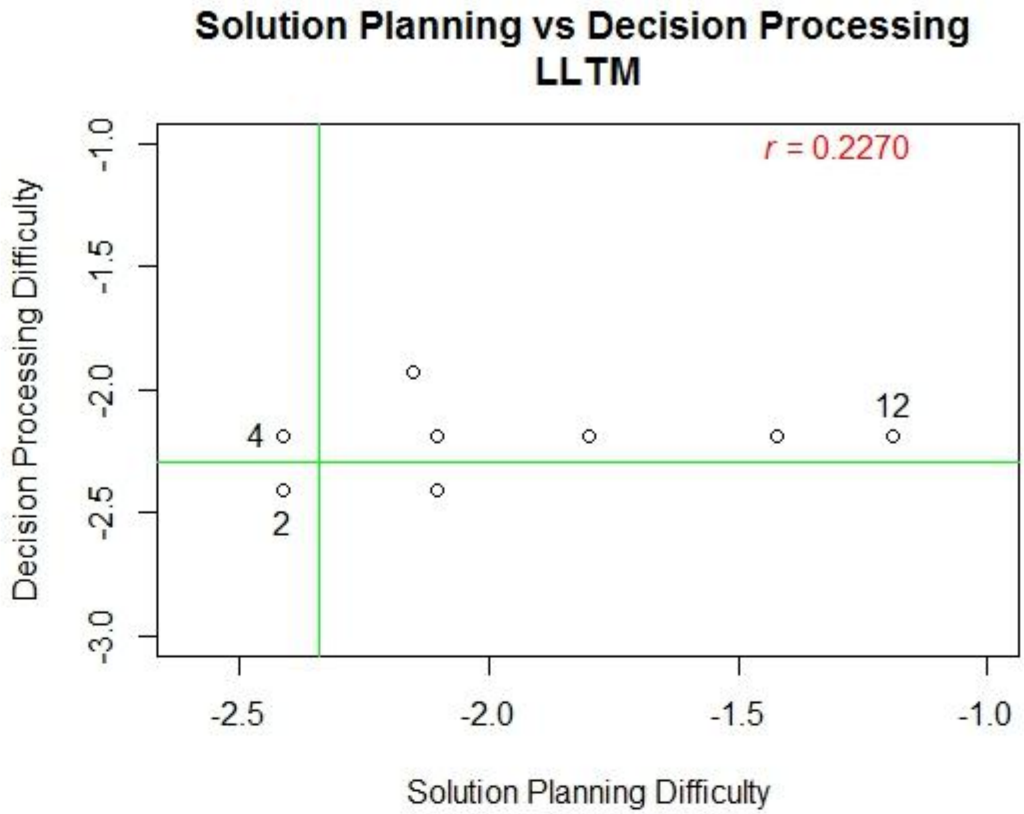


Figure D 8. *Solution Planning component difficulty compared to Solution Execution component difficulty for the LLTM.*



\

Figure D 9. *Solution Planning component difficulty compared to Decision Processing component difficulty for the LLTM.*

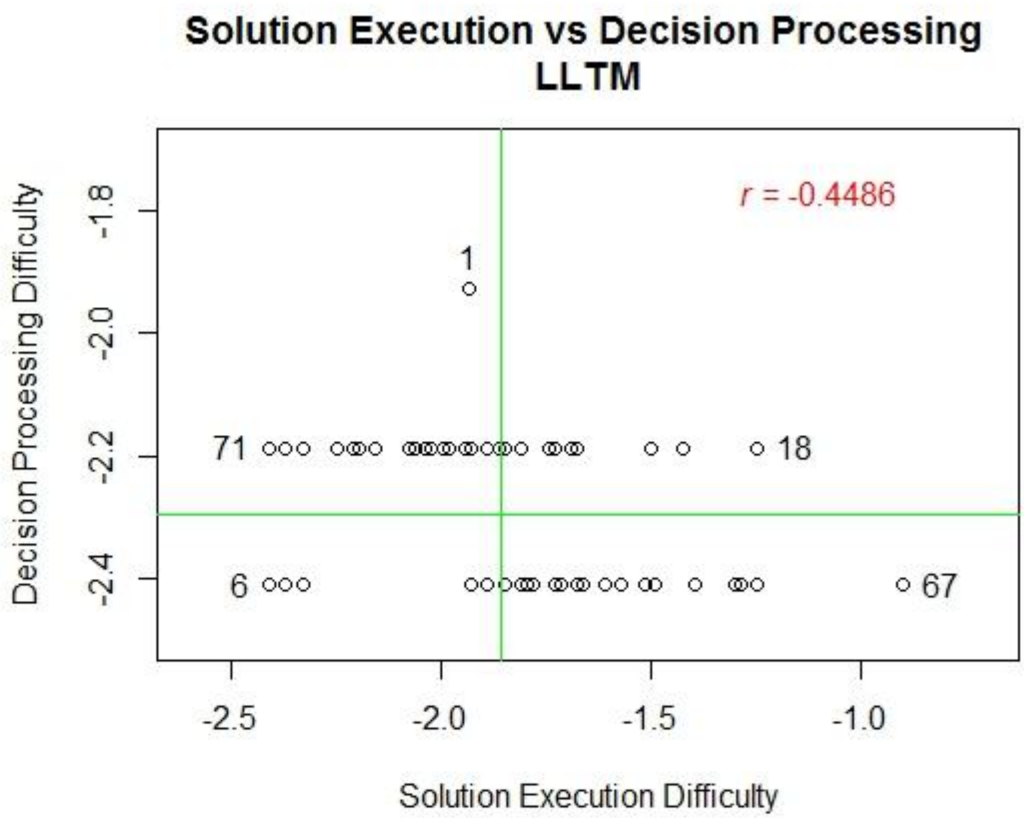


Figure D 10. *Solution Execution component difficulty compared to Decision Processing component difficulty for the LLTM.*

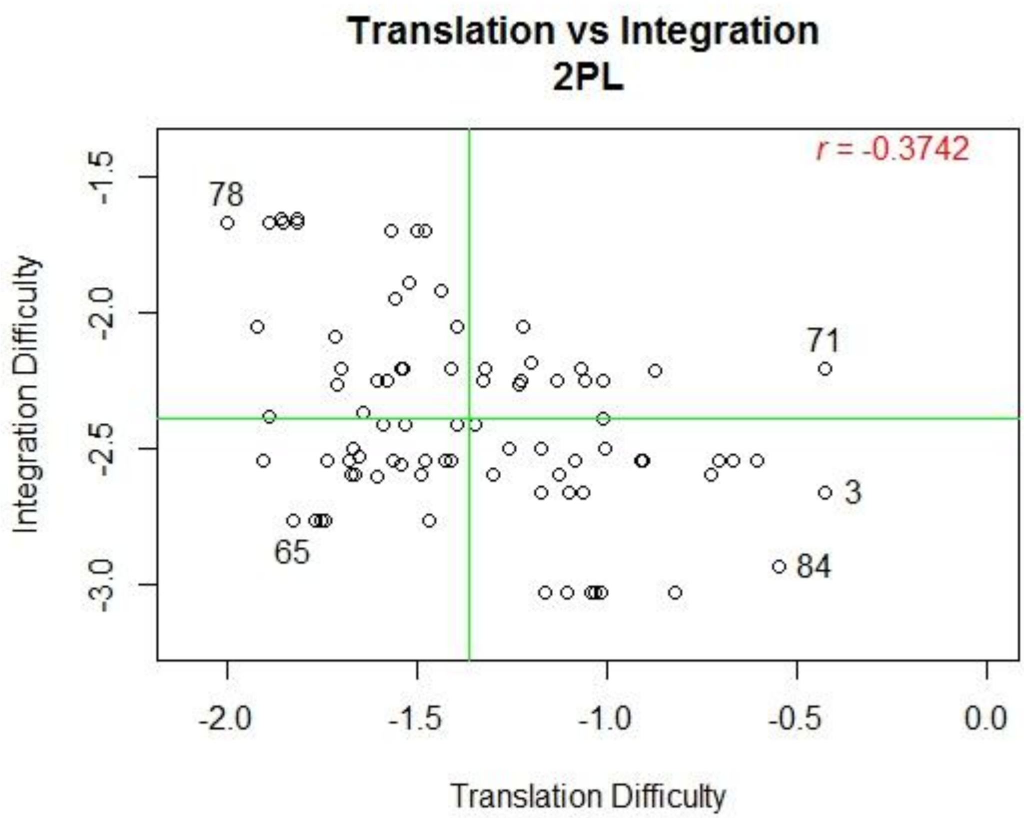


Figure D 11. *Translation component difficulty compared to Integration component difficulty for 2PL regression.*

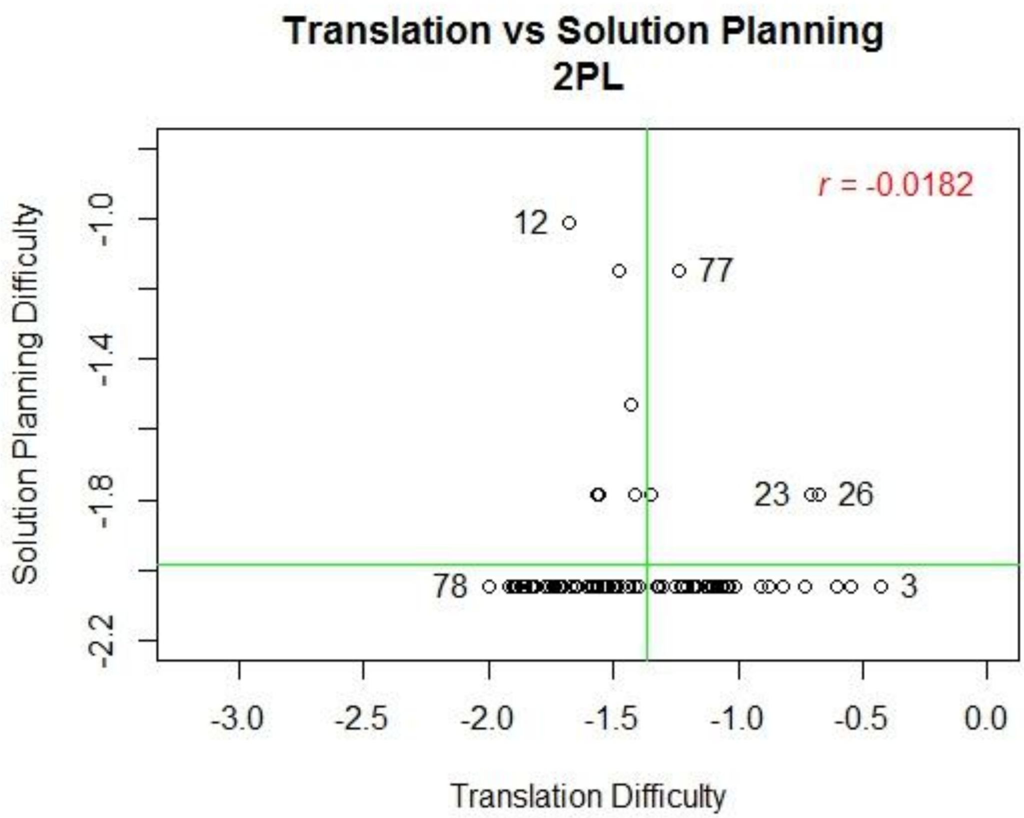


Figure D 12. *Translation component difficulty compared to Solution Planning component difficulty for 2PL regression.*

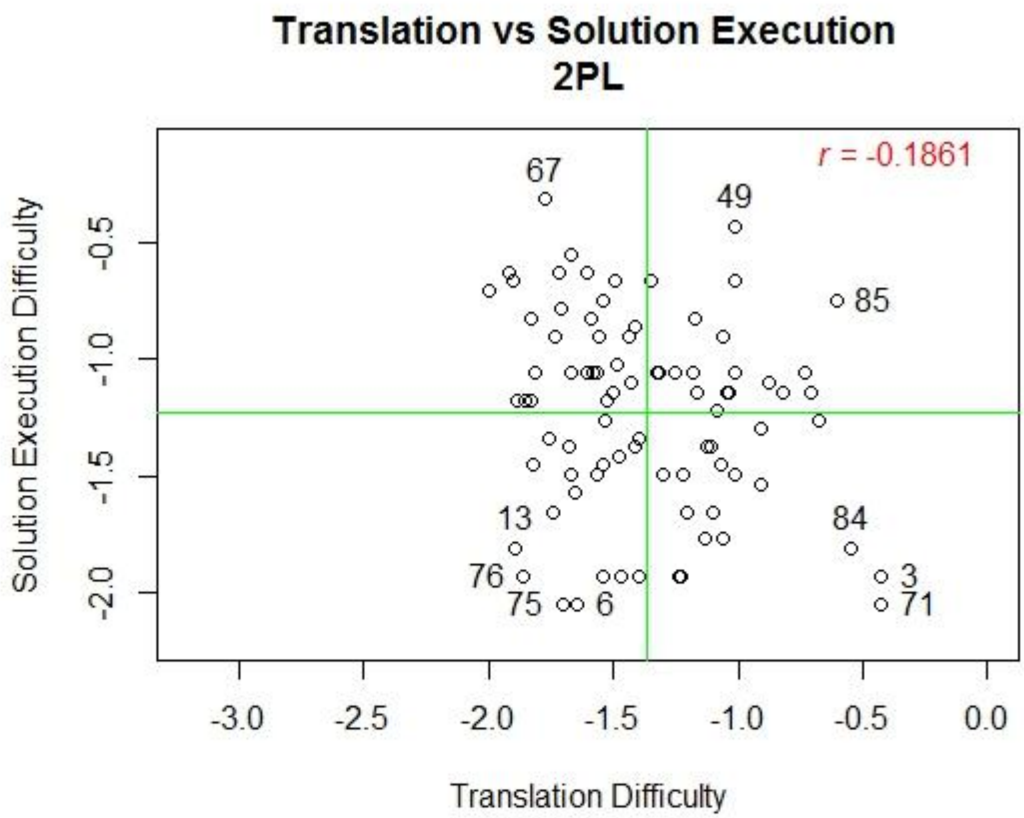


Figure D 13. *Translation component difficulty compared to Solution Execution component difficulty for 2PL regression.*

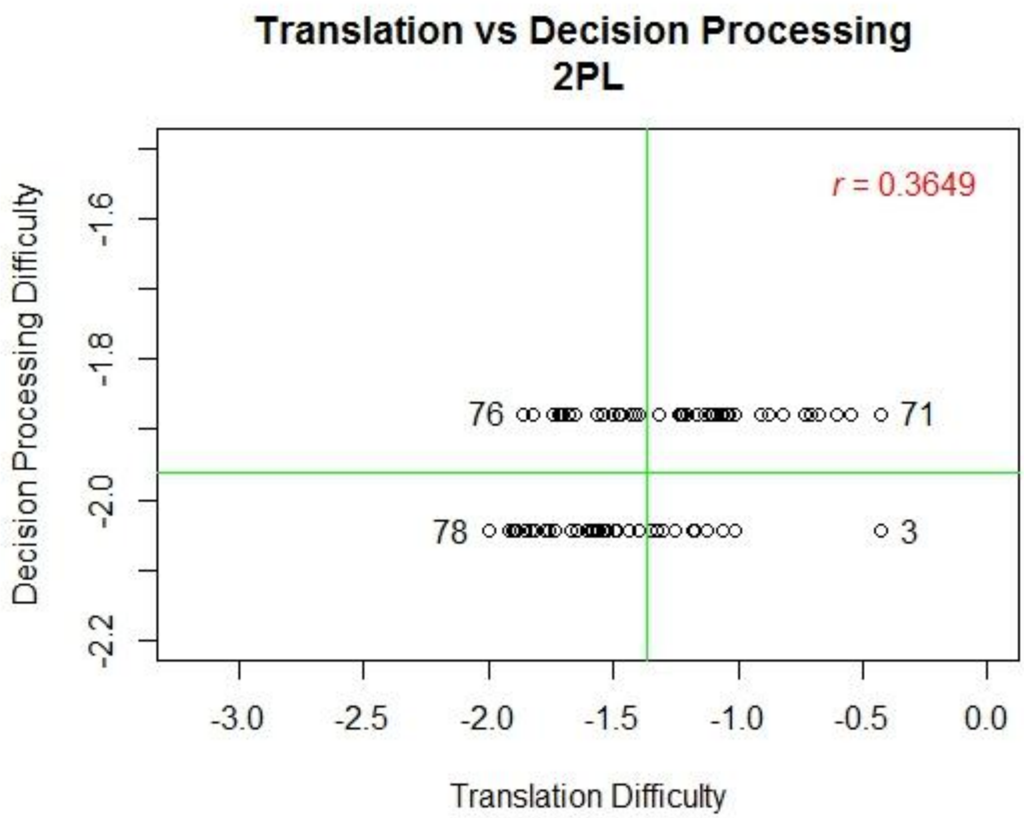


Figure D 14. *Translation component difficulty compared to Decision Processing component difficulty for 2PL regression.*

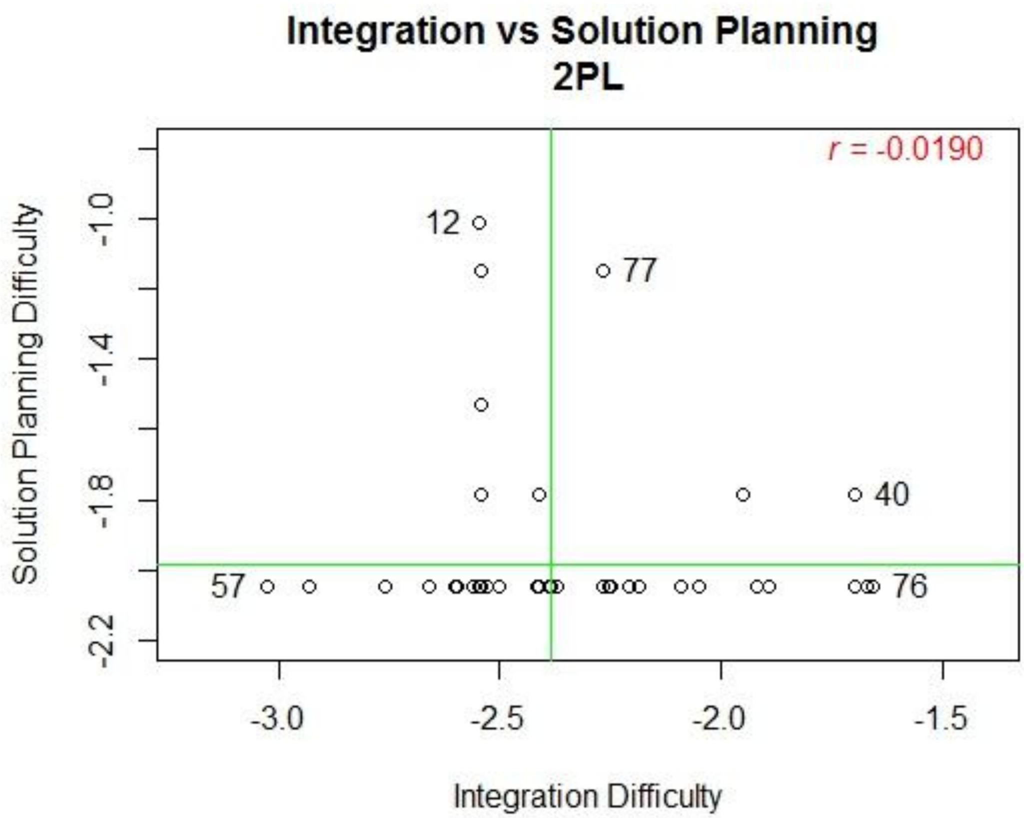


Figure D 15. *Integration component difficulty compared to Solution Planning component difficulty for 2PL regression.*

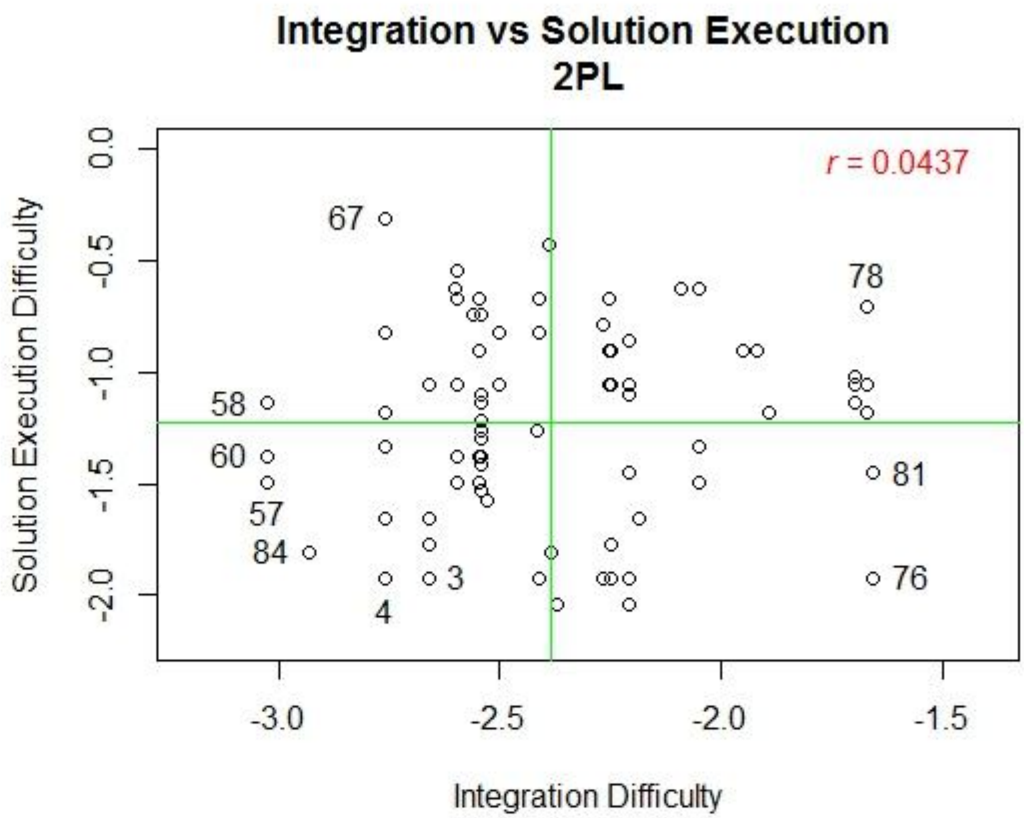


Figure D 16. *Integration component difficulty compared to Solution Execution component difficulty for 2PL regression.*

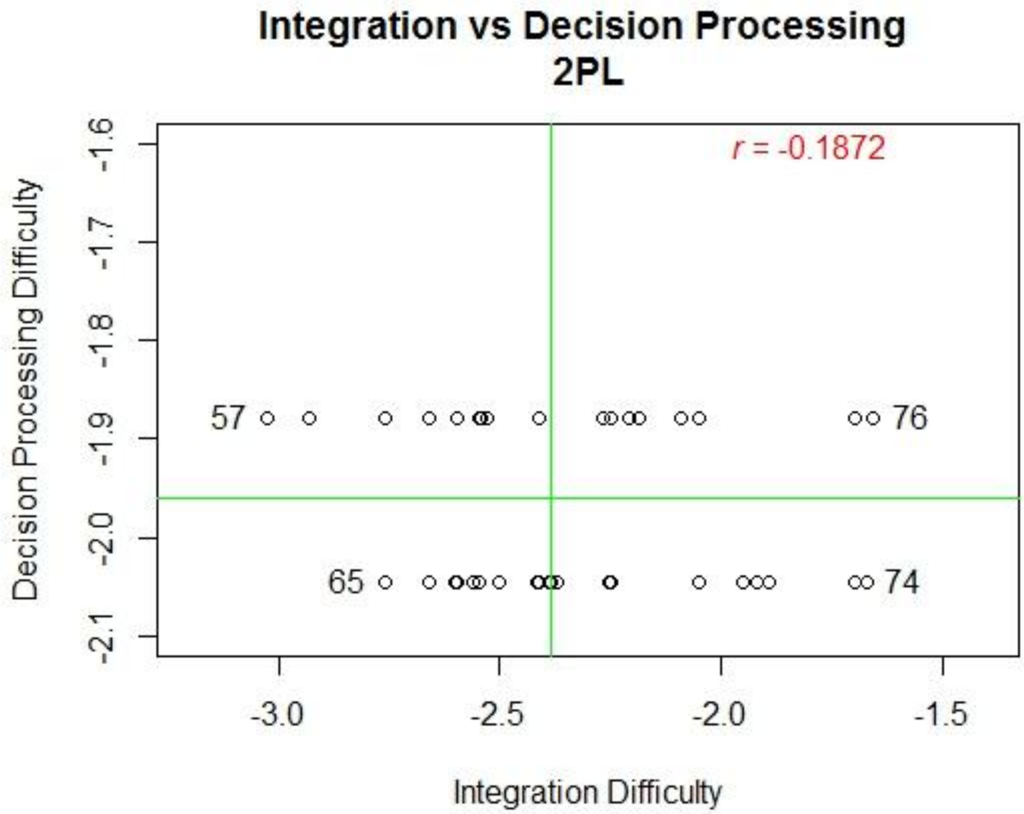
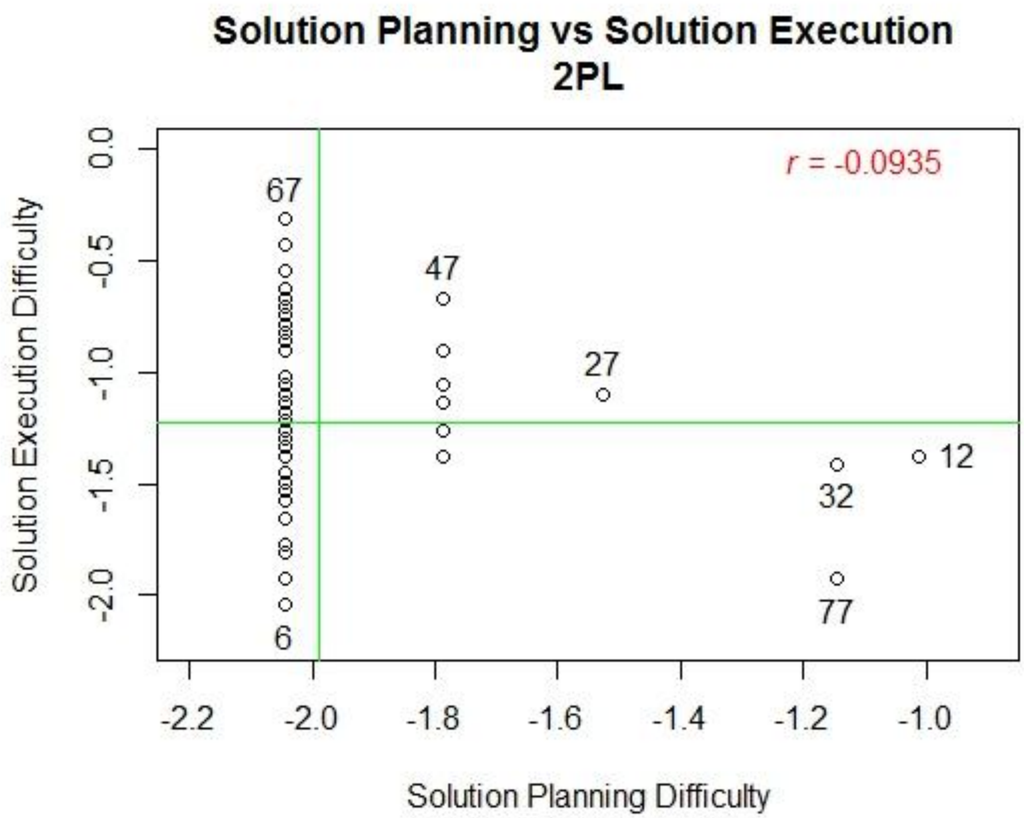


Figure D 17. *Integration component difficulty compared to Decision Processing component difficulty for 2PL regression.*



\

Figure D 18. *Solution Planning component difficulty compared to Solution Execution component difficulty for 2PL regression.*

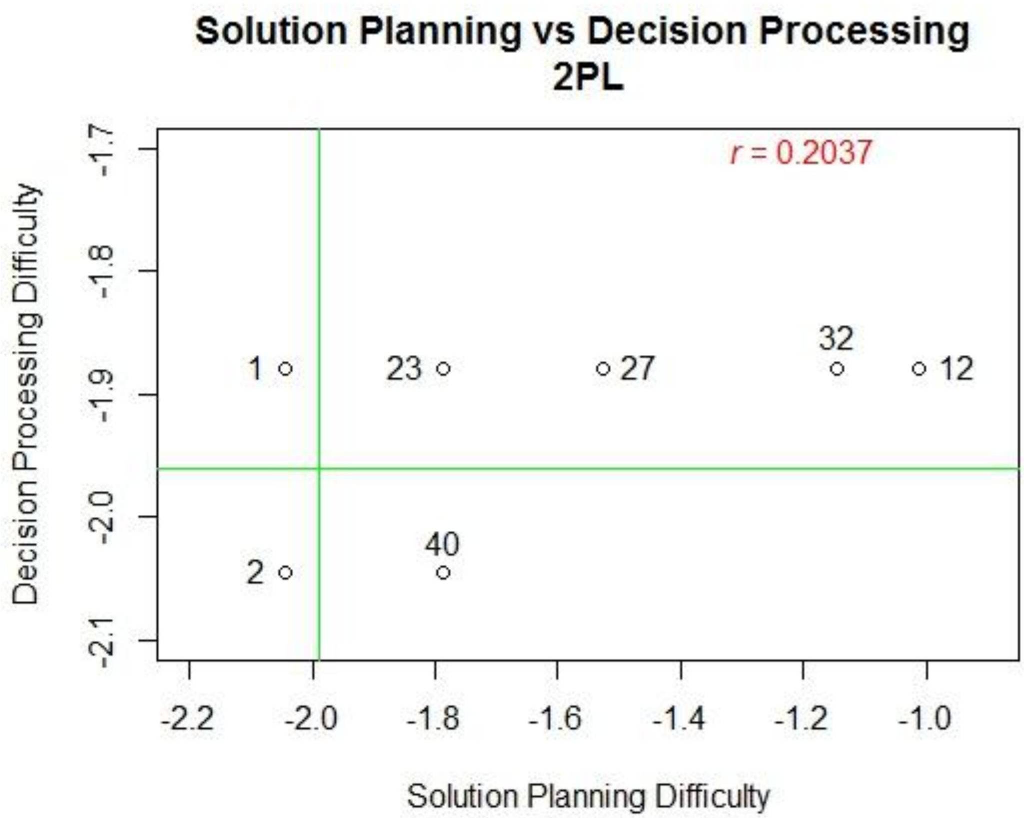


Figure D 19. *Solution Planning component difficulty compared to Decision Processing component difficulty for 2PL regression.*

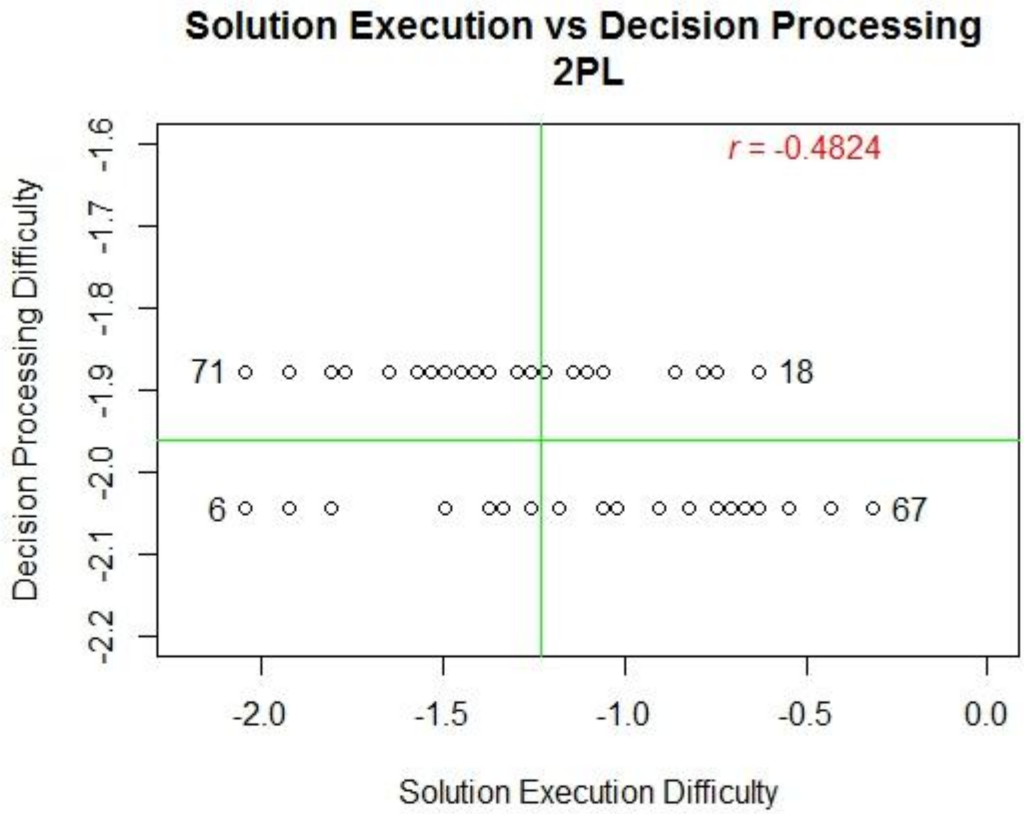


Figure D 20. *Solution Planning component difficulty compared to Decision Processing component difficulty for 2PL regression.*

REFERENCES

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131*(1), 30-60.

Ashcraft, M. H., & Battaglia, J. (1978). Cognitive Arithmetic: Evidence for retrieval and decision processes in mental addition. *Journal of Experimental Psychology: Human Learning and Memory, 4*(5), 527-538.

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. Journal of Experimental Psychology: General, 130, 224-237.

Ashcraft, M., & Krause, J. (2007). Working memory, math performance, and math anxiety. *Psychonomic Bulletin & Review, 14*(2), 243-248.

Bejar, I. I. (2010). Recent developments and prospects in item generation. In Embretson, S. E. (Ed.), *Measuring psychological constructs: Advances in model-based approaches*. Washington, D.C.: American Psychological Association Books.

Bennett, R. E. (2010). Cognitively Based Assessment *of*, *for*, and *as* Learning (CBAL): A preliminary theory of action for summative and formative assessment. *Measurement: Interdisciplinary Research & Perspectives, 8*(2-3), 70-91.

Birenbaum, M., Tatsuoka, K. K., & Gurtvirtz, Y. (1992). Effects of response format on diagnostic assessment of scholastic achievement. *Applied Psychological Measurement, 16*(4), 353-3663.

Conway, A., Cowan, N, Bunting, M. F., Therriault, D. J., & Minkoff, S. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163-183.

Cooney, J. B., & Swanson, H. L. (1990). Individual differences in memory for mathematical story problems: Memory span and problem perception. *Journal of Educational Psychology, 82*(3), 570-577.

Covington, M. A. (2007). CPIDR 3 User Manual. CASPR Research Report 2007-03. Artificial Intelligence Center, University of Georgia: Athens, Georgia.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavior and Brain Sciences*, *24*, 87-185.

Daniel, R. C. & Embretson, S. E. (2010). Designing cognitive complexity in mathematical problem solving items. *Applied Psychological Measurement*, pp. 348-364.

Dark, V. J., & Benbow, C. P. (1990). Enhanced problem translation and short-term memory: Components of mathematical talent. *Journal of Educational Psychology*, *82*(3), 420-429.

Dark, V. J., & Benbow, C. P. (1991). Differential enhancement of working memory with mathematical versus verbal precocity. *Journal of Educational Psychology, 83*(1), 48-60.

Deshler, D. D., & Lenz, B. K. (1989). The strategies instructional approach. *International Journal of Disability, Development and Education*, *3*(2), 15-23.

Embretson, S. E. (1984). A general latent trait model for response processes. *Psychometrika*, *49*(2), 175-186.

Embretson, S. E. (1999). Generating items during testing: Psychometric issues and models. *Psychometrika, 64*(4), 407-433.

Embretson, S. E. (2010). Cognitive design systems: A structural modeling approach to developing ability tests. In Embretson, S.E. (Ed.), *Measuring psychological constructs with model-based approaches.* Washington, D.C.: American Psychological Association Books, 247-274.

Embretson, S. E. & Daniel, R. C. (2008). Understanding and quantifying cognitive complexity level in mathematical problem solving items. *Psychology Science, 50,* 328- 344.

Embretson, S., Fultz, J., & Dayl, N. (1989). The influence of paragraph comprehension components on test validity. New York, NY, England: Praeger Publishers.

Embretson, S. E., & Reise, S. P. (2000). Item response theory for psychologists. Mahwah, NJ: Erlbaum.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19-23.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309-331.

Flyod, R., Evans, J., & McGrew, K. (2003). Relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools, 40*(2), 155-171.

Geary, D. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin, 114*(2), 345-362.

Gierl, M. J., Leighton, J. P., & Hunka, S. M. (2000). An NCME instructional module on exploring the logic of Tatsuoka's Rule-Space Model for test development and analysis. *Educational Measurement: Issues and Practice, 19*, 34-44.

Gorin, J. S. (2005). Manipulating processing difficulty of reading comprehension questions: The feasibility of verbal item generation. *Journal of Educational Measurement*, *42*(4), 351-373.

Gorin, J. & Embretson, S. E.. (2006). Item difficulty modeling of paragraph comprehension items. *Applied Psychological Measurement, 30*, 394-411.

Haladyna, T. M., Downing, S. M. (1988, April). Functional Distractors: Implications for Test-Item Writing and Test Design. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, Louisiana.

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *Journal of Educational Psychology, 91*(4), 684-689.

Hitch, G. J., Towse, J. N., & Hutton, U. (2001). What limits children's working memory span? Theoretical accounts and applications for scholastic development. *Journal of Experimental Psychology: General, 130*(2), 184-198.

Jones, W. P., & Anderson, J. R. (1987). Short- and long-term memory retrieval: A comparison of the effects of information load and relatedness. *Journal of Experimental Psychology: General, 116*(2), 137-153.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363-394.

Landauer, T.K. (1998) Learning and representing verbal meaning: Latent semantic analysis theory. *Current Directions in Psychological Science, 7*(5), 161-164.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2&3), 259-284.

Lutz, M.E., Embretson, S.E., & Poggio, J. (2010, May). *Assessing the components of mathematics achievement* [Poster]. Presented at the meeting of the American Educational Research Association, Denver, CO.

MacDonald, G. & Kromrey, J. (2011). Linear Logistic Test Model: Using SAS to simulate the decomposition of item difficulty by algorithm, sample size, cognitive component and time to convergence. Proceedings of the Southeast SAS Users Group, Alexandria, VA.

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). Common Core State Standards. Washington, D.C. Retrieved from http://www.corestandards.org/.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 411-421.

PARCC. (July, 2012). *Career and College Readiness for the Next Generation.* Retrieved from http://www.parcconline.org/about-parcc.

PARCC. (2012). 3-8 Assessments. Retrieved from http://www.parcconline.org/3-8-assessments.

Perin, D. (2011). Facilitating student learning through contextualization. *Community College Research Center, Brief 53*, 1-4.

Race To The Top. (2012, August 10). *Awards*. Retrieved from http://www2.ed.gov/programs/racetothetop-assessment/awards.html

Schacter, D., Wagner, A. & Buckner, R. (2000). Memory Systems of 1999. In E. Tulving (Ed.) & F. Craik (Ed.), *The Oxford Handbook of Memory*. New York, NY: Oxford University Press.

Smarter Balanced Assessment Consortium. (n.d.). *Executive Summary of Smarter Balanced Race to the top Assessment Application.* Retrieved from http://www.smarterbalanced.org/about/.

Smarter Balanced Assessment Consortium. (n.d.). *Introduction to Smarter Balanced item and performance task development.* Retrieved from http://www.smarterbalanced.org/smarter-balanced-assessments/item-writing-and-review/.

Smarter Balanced Assessment Consortium. (n.d.). *Smarter Balanced mathematics item specifications grades 6-8.* Retrieved from http://www.smarterbalanced.org/smarter-balanced-assessments/item-writing-and-review/.

Smarter Balanced Assessment Consortium. (2012). *Performance Tasks Specifications*. Retrieved from http://www.smarterbalanced.org.

Swanson, H., & Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology, 96*(3), 471-491.

Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology*, *80*(4), 424-436.

Turner, A., & Greene, E. (1977). The Construction and Use of a Propositional Text Base. (Technical Report No. 63). Boulder, Colorado: Institute for the Study of Intellectual Behavior.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*(1), 104-132.

van Garderen, D., & Montague, M. (2003). Visual-spatial representations and mathematical problem solving. *Learning Disabilities & Research*, *18*, 246-254.

Wakefield, J. A. (1958). Does the fifth choice strengthen a test item? *Public Personnel Review, 19*, 44-48.

Widaman, K. F., Geary, D. C., Cormier, P., & Little, T. D. (1989). A componential model for mental addition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15(5), 898-919.

Zentall, S. (1990). Fact-retrieval automatization and math problem solving by learning disabled, attention-disordered, and normal adolescents. *Journal of Educational Psychology*, *82*(4), 856-865.

Zimowski, Muraki, Mislevy, & Bock. (2003). *BILOG-MG 3: Item analysis and test scoring with binary logistic models.* Chicago, IL: Scientific Software. [Computer software].